



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ,
ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ**

**Ανάπτυξη μεθόδων βαθιάς μάθησης για
την αναγνώριση βιομηχανικών
προϊόντων**

*Διπλωματική εργασία που εκπονήθηκε στο Μεταπτυχιακό Πρόγραμμα Σπουδών στην
Εφαρμοσμένη Πληροφορική για τη χορήγηση του Μεταπτυχιακού Διπλώματος*

ΑΛΕΞΑΝΔΡΟΣ ΚΑΠΑΝΔΕΛΗΣ
Α.Μ.: 228

Επιβλέπων Καθηγητής
Δημήτριος Βαρσάμης

Σέρρες, Νοέμβριος 2023

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ ΓΙΑ ΤΗΝ
ΑΝΑΓΝΩΡΙΣΗ ΒΙΟΜΗΧΑΝΙΚΩΝ ΠΡΟΪΟΝΤΩΝ

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο υπογράφων Αλέξανδρος Καπανδέλης του Μιχαήλ, με αριθμό μητρώου 228, φοιτητής του τμήματος Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών της σχολής Μηχανικών του Διεθνούς Πανεπιστημίου Ελλάδος, δηλώνω υπεύθυνα ότι:

«Αυτή η διπλωματική εργασία υποβάλλεται από τον συγγραφέα της στα πλαίσια εκπλήρωσης των απαιτήσεων του Μεταπτυχιακού Προγράμματος Σπουδών στην Εφαρμοσμένη Πληροφορική του τμήματος Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών της Σχολής Μηχανικών του Διεθνούς Πανεπιστημίου Ελλάδος.

Υπεύθυνα δηλώνεται ότι, η συγκεκριμένη διπλωματική εργασία έχει συγγραφεί από τον Υπογράφοντα και δεν έχει υποβληθεί ούτε έχει αξιολογηθεί στο πλαίσιο άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό.»

Σέρρες, 15.11.2023

Ο Δηλών

ΠΕΡΙΛΗΨΗ

Η εξέλιξη της τεχνολογίας στις μέρες μας έχει επιφέρει σημαντική πρόοδο και βελτίωση σε κάθε τομέα της βιομηχανίας. Εξειδικευμένα - συνεχώς εξελισσόμενα - τεχνολογικά μέσα ενσωματώνονται καθημερινά σε βιομηχανικές μονάδες, προκειμένου να διευκολύνουν την υλοποίηση της εκάστοτε παραγωγής, να οργανώσουν ένα σύνολο εργασιών που είναι απαραίτητο να εκτελεστούν στα πλαίσια ενός κύκλου εργασιών μιας μονάδας, να βελτιστοποιήσουν τον τομέα ανάπτυξης και του marketing της βιομηχανίας, αλλά και να πετύχουν την αμεσότερη και καλύτερη επικοινωνία των διαφόρων τμημάτων είτε στο επίπεδο της μονάδας είτε μεταξύ των μονάδων που διαθέτει εκάστοτε βιομηχανία (ακόμα και για την επικοινωνία παραρτημάτων σε παγκόσμιο επίπεδο). Έτσι, παρατηρείται η συνεχής ενσωμάτωση νέων τεχνολογιών σε βιομηχανικές μονάδες, είτε μέσω υλικολογισμικού είτε μέσω νέων – βελτιωμένων λογισμικών συστημάτων σε κάθε τμήμα μιας δομημένης βιομηχανίας.

Στην εργασία αυτή ερευνούμε την Οπτική Αναγνώριση Βιομηχανικών Προϊόντων ακολουθώντας δύο προσεγγίσεις που περιγράφονται παρακάτω. Για την αξιολόγηση κάθε προσέγγισης, χρησιμοποιούμε σύνολα δεδομένων και πειραματικές ρυθμίσεις που ακολουθούνται από τη βιβλιογραφία. Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση και η σύγκριση τεχνικών υπολογιστικής όρασης και τεχνικών βαθιάς μάθησης. Είναι δύο εντελώς διαφορετικές προσεγγίσεις του θέματος, καθεμία με τα δικά της πλεονεκτήματα.

Στην πρώτη προσέγγιση εξάγουμε χαρακτηριστικά από τις εικόνες και ταξινομούμε ένα σύνολο εικόνων με βάση τα χαρακτηριστικά αυτά. Χρησιμοποιούμε χαρακτηριστικά που είναι αμετάβλητα ως προς την κλίμακα, ως προς την περιστροφή καθώς και μία μέθοδο ταχείας εξαγωγής. Δεδομένου ότι ολόκληρη η διαδικασία ανίχνευσης είναι ένα ενιαίο δίκτυο, μπορεί να βελτιστοποιηθεί άμεσα από άκρο σε άκρο. Για την αξιολόγηση των μεθόδων και την εκτέλεση των πειραμάτων χρησιμοποιούμε το σύνολο δεδομένων που συλλέξαμε με τη χρήση μη επανδρωμένου αεροχήματος (Drone).

Στην δεύτερη προσέγγιση κάνουμε χρήση Συνελκτικών Νευρωνικών Δικτύων, τα οποία εκπαιδεύονται στο σύνολο δεδομένων. Χρησιμοποιούμε είτε εκπαιδευμένα εκ

των προτέρων νευρωνικά δίκτυα είτε όχι για την παραγωγή των μοντέλων. Για την αξιολόγηση των μεθόδων και την εκτέλεση των πειραμάτων χρησιμοποιούμε το σύνολο δεδομένων.

Λέξεις Κλειδιά:

Συνελκτικά Νευρωνικά Δίκτυα, Μηχανική Μάθηση, Βαθιά Μάθηση, Μεταφορά Μάθησης, Όραση Υπολογιστών, Αναγνώριση Φαγητού, Αναλλοίωτος σε κλίμακα Μετασχηματισμός, Επιτυχαχυνόμενος Μετασχηματισμός, Μη επανδρωμένα αεροχήματα, Τεχνολογία LiDAR, Έλεγχος Αποθήκης - Monitoring, On-the-fly Αναγνώριση

ABSTRACT

The introduction of automation technologies in product warehouses is considered essential today, for the most effective and smoothest operation and organization of storage inventory control and monitoring. The significance of integrating automation technologies is particularly significant, especially in the modern era of globalization, as numerous global industrial companies with a wide range of products need to control and organize their distribution warehouses in the most modern and beneficial way possible. Specifically, replacing the extensive human factor, required in previous years for organizing distribution warehouses, is rapidly increasing, as technological means come into consideration. As a result, the risk of human error is minimized, control is carried out promptly and securely, and the cost for companies is significantly reduced, focusing primarily on the acquisition and maintenance of technological tools.

A form of technology sought to be integrated into the organization and overall efficient operation of warehouses is considered to be the use of Unmanned Aerial Vehicles (UAVs or Drones). Specifically, with the technologies with which a drone can be equipped and in combination with suitable software, the direct detection and identification of the inventory of a product warehouse can be achieved. The aforementioned process provides the necessary technological support in the field of logistics for each company and assists in the best possible way the respective product storage department and the employees working in it, keeping them constantly informed about the inventory status and the overall productivity of the distribution warehouse.

In this work, we investigate on Optical Industrial item Recognition in a Warehouse following two approaches that are described below. For the evaluation of each approach, we use datasets and configuration used in bibliography. The aim of my thesis is to present and compare computer vision and deep learning techniques. They are very different approaches, each one with its own advantages.

In the first approach, we extract features from the images and classify a dataset of images according to these features. We use scale and rotation invariant features as well as a method of fast features extraction. As the whole procedure of classification is a

unified network, it can be end-to-end optimized. For the evaluation of the methods and the experiments we use our own dataset collected from a local warehouse.

In the second approach, we use Convolutional Neural Networks that are trained on the same dataset. We use either pre-trained neural networks or not for the model production. For the methods evaluation and the experiments, we use the same dataset.

Keywords:

Convolutional Neural Networks, Machine Learning, Deep Learning, Transfer Learning, Computer Vision, Industrial Item Recognition, SIFT, SURF, Drones, LiDAR technology, Inventory monitoring, On-the-fly detection

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα καταρχήν να ευχαριστήσω τον κ. Δημήτριο Βαρσάμη για την επίβλεψη αυτής της διπλωματικής εργασίας. Επιπλέον ατέρμονες είναι οι ευχαριστίες προς τους γονείς μου για την αγάπη και την υποστήριξη που μου προσέφεραν καθ' όλη την ακαδημαϊκή μου περίοδο. Ειδικότερα, τους ευχαριστώ για την υπευθυνότητα και την δημιουργικότητα, αξίες με τις οποίες με γαλούχησαν.

Τέλος, θα ήθελα να ευχαριστήσω όλους τους φίλους και τους ανθρώπους που μου στάθηκαν στις πιο δύσκολες στιγμές της ζωής μου. Ποτέ δε θα ξεχάσω την διαρκή ενθάρρυνση τους καθώς μοιραζόμουν μαζί τους τα όνειρα και φιλοδοξίες.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ	11
ΚΕΦΑΛΑΙΟ 1 ^ο	16
Τι είναι η Βαθιά Μάθηση	16
Πλεονεκτήματα Deep Learning και 5 λόγοι χρήσης	19
ΚΕΦΑΛΑΙΟ 2 ^ο	20
Πώς λειτουργεί Το deep learning ;	20
Γιατί Βαθιά Μάθηση;	22
ΚΕΦΑΛΑΙΟ 3 ^ο	23
Άνοδος της βαθιάς μάθησης	23
Δράση του Deep Learning	24
Παραδείγματα βαθιάς μάθησης	25
ΚΕΦΑΛΑΙΟ 5 ^ο	27
Βαθιά μάθηση έναντι μηχανικής μάθησης	27
Μέθοδοι Βαθιάς Μάθησης	28
ΚΕΦΑΛΑΙΟ 6 ^ο	30
Προοπτικές σταδιοδρομίας Deep Learning	30
Πλεονεκτήματα Της Βαθιάς Μάθησης.....	30
Αλγόριθμοι βαθιάς μάθησης.....	31
Εμπόδια στην εφαρμογή της βαθιάς μάθησης.....	31
ΚΕΦΑΛΑΙΟ 7 ^ο	33
Απαιτήσεις υλικού Deep Learning	33
Deep Learning και IBM.....	33
Περιορισμοί και προκλήσεις.....	33
Κεφάλαιο 8 ^ο	36
Επανάσταση Βαθιάς Μάθησης.....	36
Κριτική και σχόλιο.....	39
Σφάλματα.....	39
Κυβερνοαπειλή.....	40
Εξάρτηση από τον άνθρωπο	41
ΚΕΦΑΛΑΙΟ 9 ^ο	43
Μηχανική Μάθηση	43
Ορισμός.....	43
Μάθηση με Επίβλεψη	43
Μάθηση χωρίς Επίβλεψη.....	45
Νευρωνικά Δίκτυα	45
Optimizers.....	48
Γραμμική Ταξινόμηση	51
Λογιστική Παλινδρόμηση.....	54
Μεταφορά Μάθησης.....	55
ΚΕΦΑΛΑΙΟ 10 ^ο	58
Συνελκτικά Νευρωνικά Δίκτυα: Εισαγωγή	58
Στοιχεία Συνελκτικού Νευρωνικού Δικτύου	58
ΚΕΦΑΛΑΙΟ 11 ^ο	62
Προτεινόμενο Σύστημα	62

Αναλλοίωτος σε Κλίμακα Μετασχηματισμός	62
Επιταχυνόμενος Μετασχηματισμός.....	66
Τσάντα Οπτικών Λέξεων.....	69
Αρχιτεκτονική Συνελκτικού Δικτύου	72
ΚΕΦΑΛΑΙΟ 12°	77
Διεξαγωγή Πειραμάτων	77
Πειραματικές Ρυθμίσεις.....	78
Αποτελέσματα πειραμάτων	78
ΚΕΦΑΛΑΙΟ 13°	80
Συμπεράσματα και μελλοντική εργασία.....	80
ΒΙΒΛΙΟΓΡΑΦΙΑ	82

ΕΙΣΑΓΩΓΗ

Το τμήμα logistics μίας μονάδας επιτελεί σημαντικό ρόλο στη παγκόσμια οικονομία σήμερα. Αποτελεί έναν τομέα της βιομηχανίας, στον οποίο απαιτείται άμεση και έγκαιρη ενημέρωση (OTA updates), καθώς και ακρίβεια στην κατάσταση των αποθεμάτων της, προκειμένου οι βιομηχανίες να έχουν μία πλήρη εικόνα της κατάστασης ανά πάσα στιγμή. Οι βιομηχανίες σήμερα επιβάλλεται να έχουν μία σωστή εικόνα του τμήματος logistics και βασίζονται στην αξιόπιστη και ανέπαφη πληροφορία που αυτά παρέχουν προκειμένου να βελτιώσουν στην οικονομική τους αξία. Κρίσιμο λοιπόν θεωρείται, τα τμήματα logistics της εκάστοτε βιομηχανικής μονάδας να παρέχουν ανά πάσα στιγμή έγκυρες πληροφορίες σχετικά με το απόθεμα που υπάρχει σε κάθε αποθήκη οπουδήποτε αυτή και αν βρίσκεται ανά την υφήλιο. Επομένως, κρίνεται απαραίτητη η αυτοματοποίηση του μηχανισμού ανίχνευσης και αναγνώρισης αποθέματος προϊόντων, προκειμένου να παρέχεται στον μικρότερο δυνατό χρόνο και χωρίς την ύπαρξη λάθους - το οποίο συμβαίνει λόγω του κινδύνου λάθους που περιέχει ο ανθρώπινος παράγοντας - η πραγματική κατάσταση των προϊόντων που έχει στη διάθεσή της μία βιομηχανία, με αποτέλεσμα να βελτιώνεται ο χρόνος εξυπηρέτησης του αγοραστικού κοινού και να ελαχιστοποιείται το κόστος της παραγωγής.

Σημαντική βελτίωση πιστεύεται ότι θα φέρει στον τομέα αυτό η χρήση μη επανδρωμένων αεροχημάτων (UAV), κοινώς η χρήση “Drone”, τα οποία με τα κατάλληλα υλικοτεχνικά και λογισμικά συστήματα, θα μπορούν να καταγράφουν με αυξημένη αξιοπιστία την πραγματική κατάσταση του αποθέματος μίας αποθήκης προϊόντων. Ήδη η εξέλιξη των συστημάτων αυτών, τα έχει κάνει ικανά να κινούνται σε εσωτερικούς και εξωτερικούς χώρους με σχετική ή και πλήρη αυτονομία και χωρίς την ανθρώπινη παρέμβαση. Κρίνεται επομένως σημαντικό να εξευρεθεί ένα κατάλληλο λογισμικό σύστημα, το οποίο, αξιοποιώντας τα δεδομένα που δίνει το μη επανδρωμένο αερόχημα και, με απόλυτο συγχρονισμό, να ολοκληρώνει με απόλυτη ακρίβεια την πλήρη καταγραφή των ποσοτήτων των προϊόντων που υπάρχουν σε κάθε αποθήκη μίας βιομηχανίας. Η εξεύρεση ενός τέτοιου αλγόριθμου εξαρτάται από τις τεχνολογίες που θα επιλεγθούν και με τις οποίες θα εξοπλίζεται ένα drone. Αναλόγως των ανωτέρω τεχνολογιών που θα χρησιμοποιηθούν, προκύπτουν διαφορετικές προσεγγίσεις στην υλοποίηση ενός αποδοτικού αλγόριθμου που θα ολοκληρώνει την ανωτέρω διεργασία με τον βέλτιστο δυνατό τρόπο. Απαιτείται λοιπόν η προσεκτική

εξέταση και επιλογή τεχνολογικών μέσων που σε συνδυασμό με τον αλγόριθμό, θα παράγουν αποτελέσματα με ακρίβεια και με το λιγότερο κόστος υλοποίησης.

Αναφορικά με τις υφιστάμενες έρευνες που έχουν πραγματοποιηθεί στον τομέα ανίχνευσης - αναγνώρισης στοιβάς προϊόντων, η Reynaldo M. Gago, Matheus Y. A. Pereira, Guilherme A. S. Pereira επικεντρώνεται στην ανάπτυξη και αξιολόγηση ενός εναέριου ρομποτικού συστήματος (UAV) για την «έξυπνη» απογραφή συσσωρευμένων υλικών - αποθεμάτων εντός μία αποθήκης. Συγκεκριμένα, αφορά στην καταμέτρηση στοιβών υλικών μάζας διαφορετικής πυκνότητας και σύστασης (όπως είναι οι ενώσεις φώσφορου και καλίου – αγροτικά εφόδια), τα οποία αποθηκεύονται σε σημεία - αποθετήρια διαφορετικών μεγεθών στο εσωτερικό μιας αποθήκης. Η ανωτέρω προτεινόμενη διεργασία πραγματοποιείται με τη χρήση μη επανδρωμένου αεροχήματος - UAV, το οποίο εξοπλίζεται με σύστημα ανίχνευσης LIDAR. Το μη επανδρωμένο αερόχημα (Drone) έχει προγραμματιστεί να κινείται με σχετική αυτονομία (χρήση στρατηγικής κινήσεων μέσω εντολών χειριστή) μέσα στην αποθήκη. Μέσω της χρήσης του συστήματος LIDAR, επιδιώκεται η συγκέντρωση δεδομένων σημείων νέφους (τεχνολογία PCL), ώστε στη συνέχεια, με την κατάλληλη εκ των υστέρων επεξεργασία, ο αλγόριθμος να δημιουργεί ένα τρισδιάστατο μοντέλο του περιβάλλοντος χώρου (αποθήκης και αποθετηρίων), από το οποίο θα γίνονται οι μετρήσεις των υλικών - αποθεμάτων. Με επιμέρους επεξεργασίες, οι ερευνητές κατάφεραν να εκτιμήσουν τις ποσότητες των μαζικών υλικών κάθε στοιβάς με ακρίβεια 98 % με τη χρήση δεδομένων προσομοίωσης και πραγματικών δεδομένων. Η ως άνω διεργασία, ωστόσο, προϋποθέτει να ενσωματωθεί στον αλγόριθμο μια ακριβής απεικόνιση της αποθήκης μέσω ψηφιοποίησης και εισαγωγής κατόψεων και άλλων αρχιτεκτονικών σχεδίων, ενώ κρίθηκε απαραίτητο να εγκατασταθούν στον αλγόριθμο κάποια σταθερά σημεία (εν προκειμένω τα μεταλλικά τόξα του σκελετού της αποθήκης), ώστε να μπορέσει ο αλγόριθμος να αφαιρέσει τα περιττά δεδομένα (όπως για παράδειγμα την κάτοψη του κτηρίου της αποθήκης) και να προχωρήσει σε εκτίμηση των ποσοτήτων των υλικών. Επιπροσθέτως κρίθηκε απαραίτητο να δηλώνεται για κάθε αποθετήριο ξεχωριστά η πυκνότητα εκάστου υλικού.

Οι Tucci G., Gebbia A., Conti A., Fiorini L. και Lubello C. επιχειρούν την κατά προσέγγιση καταμέτρηση συσσωρευμένων υλικών (εν προκειμένω ανακυκλώσιμα απόβλητα) τοποθετημένα σε ανοιχτή τοποθεσία μέσω της χρήσης της διαδικασίας

«Φωτογραμμετρικής Μελέτης». Η παραπάνω διαδικασία βασίζεται σε επεξεργασία δεδομένων που συγκεντρώνονται από την φωτογράφιση των προς μέτρηση υλικών με τη χρήση ενός μη επανδρωμένου αεροχήματος (UAV). Βασισμένη στη ανωτέρω διαδικασία, οι ερευνητές επιδιώκουν να δημιουργήσουν ένα τρισδιάστατο μοντέλο του αντικείμενου – στόχου και να εξάγουν τα μετρικά χαρακτηριστικά αυτού, μέσω της συγκέντρωση και επεξεργασίας των φωτογραφιών του UAV(Drone). Στόχος της μελέτης είναι η ανά τακτά χρονικά διαστήματα επίβλεψη – παρακολούθηση των συσσωρευμένων υλικών χωρίς ανθρώπινη παρέμβαση, ελαχιστοποιώντας συγχρόνως και το κόστος που απαιτούνταν με τη χρήση των παραδοσιακών τεχνικών καταμέτρησης (μέσω κλασικών εργαλείων τοπογραφίας με χρήση ολικών σταθμών ή κινηματικών συστημάτων παγκόσμιου δορυφορικού συστήματος πλοήγησης σε πραγματικό χρόνο (GNSS-RTK).

Περαιτέρω, οι Bohan Yoon, Hyeonha Kim, Geonsik Youn και J. Rhee, ασχολούνται με την αυτοματοποίηση της διαχείρισης αποθεμάτων σε μεγάλη αποθήκη προϊόντων. Συγκεκριμένα, επιχειρούν την ανίχνευση και αναγνώριση του αποθέματος διαφόρων προϊόντων με τη χρήση τεχνολογίας UAV (Drone) και τεχνολογίας QRCode (Quick Response Code). Το μη επανδρωμένο αερόχημα έχει εξοπλιστεί με QRCode Scanner, το οποίο, κατευθυνόμενο μέσα από μία προκαθορισμένη πορεία εντός της αποθήκης, επιχειρεί να αναγνωρίσει, να καταμετρήσει και να ταξινομήσει το συνολικό εμπόρευμα της αποθήκης ανιχνεύοντας τα barcodes ή QR codes που έχουν επισυναφθεί στην εκάστοτε συσκευασία. Στόχος της ανωτέρω εργασίας είναι η εξεύρεση μία μεθόδου εκτίμησης θέσης για μη επανδρωμένα αεροσκάφη και προϊόντα με βάση το μοντέλο τμηματοποίησης κώδικα QR.

Οι Yuhang Han, Qiyong Chen, Nan Pan, Xiaojue Guo και Yuqiang An προτείνουν την χρήση υψηλής ευκρίνειας φορητής συσκευής ανάγνωσης ραδιοσυχνοτήτων (RFID Reader) σε ένα UAV (Drone) για τη διενέργεια απογραφής των αποθηκών βιομηχανικών προϊόντων. Στόχος της εργασίας τους είναι η εξεύρεση του βέλτιστου αλγόριθμου για το σχεδιασμό της 3D τροχιάς των μη επανδρωμένων αεροχημάτων (UAV) σε πολύπλοκες βιομηχανικές αποθήκες, κυρίως λόγω της πολυπλοκότητας της διεργασίας του UAV για την πλήρη ανίχνευση και ανάγνωση όλων των ετικετών RFID των προϊόντων εντός της αποθήκης. Τα τελικά αποτελέσματα της έρευνας δείχνουν ότι

ο υβριδικός αλγόριθμος DE καταλληλότητας (PSO-DE) που βασίζεται στον βελτιωμένο PSO έχει μεγαλύτερη ομοιομορφία από τον αλγόριθμο DE, τον PSO και τον αλγόριθμο βελτιστοποίησης φάλαινας (WOA) και είναι πιο κατάλληλος για τον σχεδιασμό τροχιάς των drones σε πολύπλοκες βιομηχανικές αποθήκες.

Οι A. Falco, F. Narducci, A. Petrosino προτείνουν μια διαφορετική προσέγγιση, βασισμένοι στη χρήση μη επανδρωμένων εναέριων οχημάτων για τη σάρωση αποθηκών αποθήκης και συγκεχυμένου νευρωνικού δικτύου με βάση την περιοχή (R-CNN – Deep Learning) για αυτόνομες δραστηριότητες απογραφής. Αναφορικά με την απογραφή των προϊόντων, επιχειρείται η υλοποίηση ενός αλγόριθμου, ο οποίος θα αναγνωρίζει σε κάθε συσκευασία την ετικέτα που αφορά στον σειριακό αριθμό κάθε προϊόντος με σκοπό να ανιχνεύεται ο ακριβής αριθμός των προϊόντων εντός της αποθήκης, ενώ συγχρόνως θα απορρίπτονται διάφορες ετικέτες που δεν αφορούν στον ανωτέρω σκοπό. (Αλγόριθμος συσταδοποίησης).

Αντικειμενικός στόχος αυτής της έρευνας είναι η υλοποίηση ενός αλγορίθμου ο οποίος να λαμβάνει ως είσοδο ένα νέφος τρισδιάστατων σημείων και να δίνει στην έξοδό του τις ακριβείς θέσεις των στοιβών στο χώρο καθώς και το πλήθος των παλετών, τόσο ανά στοιβα όσο και συνολικά σε όλη την αποθήκη. Οι διαστάσεις της παλέτας (πλάτος, μήκος, ύψος) είναι γνωστές εκ των προτέρων και αποτελούν σταθερές του αλγορίθμου.

Οι παραπάνω έρευνες προσεγγίζουν το πρόβλημα με σχετική ακρίβεια και αποτελεσματικότητα. Ωστόσο, υπάρχουν αρκετά μειονεκτήματα ως προς την υλοποίηση του αλγορίθμου. Αυτά τα μειονεκτήματα σχετίζονται αφενός με τη χρήση πολύπλοκων υλικοτεχνικών συστημάτων (RFID Readers, QRCode Readers, πλήθος Lidar Scanner) , τα οποία ανεβάζουν το κόστος υλοποίησης και επηρεάζουν την αυτονομία πτήσης των Drone, λόγω της αύξησης του συνολικού τους - ωφέλιμο φορτίο(drone weight # drone flight autonomy). Από την άλλη πλευρά οι ανωτέρω έρευνες βασίζονται στη χρήση περίπλοκων αλγορίθμων που έχουν μεγάλο υπολογιστικό κόστος και στη χρήση πολλών περίπλοκων δεδομένων που είναι δύσκολο να διαχειριστούν. Με αποτέλεσμα να απαιτείται εκ των υστέρων επεξεργασία των δεδομένων, προκειμένου μέσω του αλγορίθμου να ολοκληρωθεί η διαδικασία αναγνώρισης του αποθέματος προϊόντων της αποθήκης.

Στόχος της έρευνας είναι η εξεύρεση του βέλτιστου αλγόριθμου ανίχνευσης και αναγνώρισης του αποθέματος μιας αποθήκης με τη χρήση μη επανδρωμένου αεροχήματος (UAV - Drone), το οποίο θα λαμβάνει στερεοσκοπικά δεδομένα με τη χρήση Lidar Scanner, για την ανίχνευση και αναγνώριση στοιβάδας προϊόντων με προκαθορισμένες διαστάσεις και σε προκαθορισμένα σημεία αποθήκευσης (αποθετήρια). Μέσω της έρευνας επιδιώκεται η εξεύρεση του βέλτιστου αλγορίθμου που θα αξιοποιεί κατάλληλα και με άμεσο τρόπο τα στερεοσκοπικά δεδομένα του LiDAR Scanner, ο όγκος των οποίων είναι κατά πολύ μικρότερος από τα αντίστοιχα δεδομένα που χρησιμοποιήθηκαν στις υφιστάμενες έρευνες, με αποτέλεσμα να είναι δυνατή η ολοκλήρωση της διαδικασίας ανίχνευσης και αναγνώρισης του αποθέματος προϊόντων on the fly, χωρίς δηλαδή την περαιτέρω αξιοποίηση των δεδομένων με περαιτέρω επεξεργασία αυτών. Με αυτόν το τρόπο επιτυγχάνεται η ελαχιστοποίηση του κόστους και χρόνου υλοποίησης της διεργασίας.

ΚΕΦΑΛΑΙΟ 1^ο

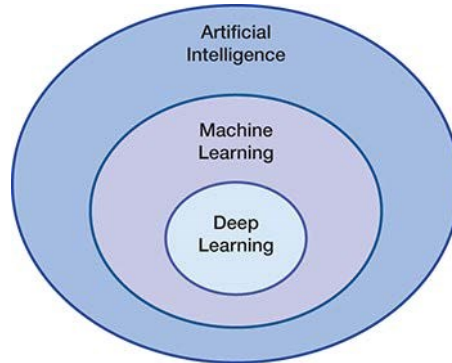
Τι είναι η Βαθιά Μάθηση

Η βαθιά μάθηση είναι μια μορφή μηχανικής μάθησης και τεχνητής νοημοσύνης που μιμείται τον τρόπο με τον οποίο οι άνθρωποι αποκτούν ορισμένους τύπους γνώσης. Η βαθιά μάθηση είναι ένα σημαντικό μέρος της επιστήμης δεδομένων συμπεριλαμβανομένων των στατιστικών και μοντέλων πρόβλεψης. Είναι εξαιρετικά χρήσιμο για επιστήμονες δεδομένων που ασχολούνται με τη συλλογή, την ανάλυση και την ερμηνεία μεγάλων ποσοτήτων δεδομένων. Η βαθιά μάθηση κάνει αυτή τη διαδικασία γρηγορότερη και ευκολότερη.

Η λέξη "deep" στη βαθιά μάθηση αναφέρεται στη χρήση πολλών επιπέδων σε ένα δίκτυο. Προηγούμενες έρευνες έδειξαν ότι τα perceptron δεν είναι ταξινομητές γενικού σκοπού, αλλά δίκτυα με μη πολυωνμικές συναρτήσεις ενεργοποίησης και κρυφά στρώματα άπειρου πλάτους. Η βαθιά μάθηση είναι μια παραλλαγή που ασχολείται με έναν αριθμό πεπερασμένων επιπέδων, επιτρέποντας πρακτική εφαρμογή και βελτιστοποίηση, διατηρώντας παράλληλα τη θεωρητική γενικότητα υπό ήπιες συνθήκες. Στη βαθιά μάθηση, τα επίπεδα επιτρέπεται επίσης να είναι διαφορετικά ως προς την αποτελεσματικότητα, την ικανότητα εκπαίδευσης και την κατανόηση, και είναι αρκετά διαφορετικά από τα συνδεδεμένα μοντέλα με βιολογικές γνώσεις, εξ ου και το «δομικό» μέρος.

Για να κατανοήσετε το deep learning, φανταστείτε ένα νήπιο του οποίου η πρώτη λέξη είναι ο σκύλος. Το νήπιο μαθαίνει τι είναι σκύλος -- και τι δεν είναι -- δείχνοντας αντικείμενα και λέγοντας τη λέξη σκύλος. Ο γονέας λέει, «Ναι, αυτός είναι σκύλος» ή, «Όχι, δεν είναι σκύλος». Καθώς το νήπιο συνεχίζει να δείχνει προς αντικείμενα, συνειδητοποιεί περισσότερο τα χαρακτηριστικά που έχουν όλα τα σκυλιά. Αυτό που κάνει το μικρό παιδί, χωρίς να το γνωρίζει, είναι να αποσαφηνίσει μια περίπλοκη αφαίρεση -- την έννοια του σκύλου -- χτίζοντας μια ιεραρχία στην οποία κάθε επίπεδο αφαίρεσης δημιουργείται με γνώση που αποκτήθηκε από το προηγούμενο στρώμα της ιεραρχίας.

Η βαθιά μάθηση είναι ένα μέρος της τεχνητής νοημοσύνης που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να έχουν προγραμματιστεί. Η βαθιά μάθηση είναι ένα επίπεδο της μηχανικής μάθησης.



Η βαθιά μάθηση σχετίζεται με τη μηχανική μάθηση που βασίζεται σε αλγόριθμους εμπνευσμένους από τα νευρωνικά δίκτυα του εγκεφάλου. Αν και ακούγεται σχεδόν σαν επιστημονική φαντασία, αποτελεί αναπόσπαστο μέρος της ανόδου της τεχνητής νοημοσύνης (AI). Η μηχανική μάθηση χρησιμοποιεί επεξεργασία δεδομένων που καθοδηγείται από αλγόριθμους, αλλά η βαθιά μάθηση προσπαθεί να μιμηθεί τον ανθρώπινο εγκέφαλο ομαδοποιώντας δεδομένα για να παράγει εκπληκτικά ακριβείς προβλέψεις.

Περιγράφει επίσης αλγόριθμους που αναλύουν δεδομένα με λογική δομή παρόμοια με τον τρόπο που οι άνθρωποι εξάγουν συμπεράσματα. Σημειώστε ότι αυτό μπορεί να επιτευχθεί τόσο με εποπτευόμενη όσο και χωρίς επίβλεψη μάθηση. Για να επιτευχθεί αυτός ο στόχος, οι εφαρμογές βαθιάς μάθησης χρησιμοποιούν μια πολυεπίπεδη αλγοριθμική δομή γνωστή ως τεχνητό νευρωνικό δίκτυο. Ο σχεδιασμός αυτού του δικτύου εμπνεύστηκε από το βιολογικό νευρωνικό δίκτυο του ανθρώπινου εγκεφάλου, με αποτέλεσμα μια διαδικασία μάθησης που είναι πιο ισχυρή από τα τυπικά μοντέλα μηχανικής μάθησης.

Σήμερα, η βαθιά μάθηση χρησιμοποιείται σε πολλούς τομείς. Στην αυτοματοποιημένη οδήγηση, για παράδειγμα, η βαθιά εκμάθηση χρησιμοποιείται για τον εντοπισμό αντικειμένων, όπως πινακίδες STOP ή πεζούς. Ο στρατός χρησιμοποιεί βαθιά μάθηση για να αναγνωρίζει αντικείμενα από δορυφόρους, π.χ. να ανακαλύψει ασφαλείς ή μη ασφαλείς ζώνες για τα στρατεύματά της. Φυσικά, η βιομηχανία ηλεκτρονικών ευρείας κατανάλωσης είναι επίσης γεμάτη βαθιά μάθηση. Για παράδειγμα, οι συσκευές

οικιακής βοήθειας όπως η Alexa της Amazon βασίζονται σε αλγόριθμους βαθιάς μάθησης για να ανταποκρίνονται στη φωνή και να μάθουν τις προτιμήσεις των ανθρώπων.

Τι θα λέγατε για ένα πιο συγκεκριμένο παράδειγμα; Φανταστείτε την εταιρεία Tesla να χρησιμοποιεί έναν αλγόριθμο βαθιάς μάθησης για τα αυτοκίνητά της για να αναγνωρίζει τα σήματα STOP. Στο πρώτο βήμα, το ANN θα προσδιορίσει τις σχετικές ιδιότητες του σήματος STOP, που ονομάζονται επίσης χαρακτηριστικά.

Τα χαρακτηριστικά μπορεί να είναι συγκεκριμένες δομές στην εισαγόμενη εικόνα, όπως σημεία, ακμές ή αντικείμενα. Ενώ ένας μηχανικός λογισμικού θα έπρεπε να επιλέξει τις σχετικές δυνατότητες σε έναν πιο παραδοσιακό αλγόριθμο Machine Learning, το ANN είναι ικανό για αυτόματη μηχανική χαρακτηριστικών. Το πρώτο στρώμα μπορεί να ανιχνεύει άκρες, το επόμενο να διακρίνει τα χρώματα και το τελευταίο κρυφό στρώμα μαθαίνει πώς να ανιχνεύει πιο περίπλοκα σχήματα. Όταν τροφοδοτούνται δεδομένα εκπαίδευσης, ένας αλγόριθμος βαθιάς μάθησης μαθαίνει τελικά από τα δικά του λάθη εάν οι προβλέψεις είναι καλές ή εάν χρειάζονται προσαρμογές.

Πλεονεκτήματα Deep Learning και 5 λόγοι χρήσης

Ένα σημαντικό πλεονέκτημα της βαθιάς μάθησης είναι ότι τα νευρωνικά της δίκτυα χρησιμοποιούνται για την αποκάλυψη κρυφών πληροφοριών και σχέσεων από δεδομένα που προηγουμένως δεν ήταν ορατά. Με πιο ισχυρά μοντέλα μηχανικής μάθησης που μπορούν να αναλύσουν μεγάλα, πολύπλοκα δεδομένα, οι εταιρείες μπορούν να βελτιώσουν τον εντοπισμό απάτης, τη διαχείριση της εφοδιαστικής αλυσίδας και την ασφάλεια στον κυβερνοχώρο αξιοποιώντας τα ακόλουθα:

Ανάλυση μη δομημένων δεδομένων: Οι αλγόριθμοι βαθιάς μάθησης μπορούν να εξετάζουν χαρακτηριστικά κειμένου αναλύοντας αναρτήσεις, ειδήσεις και έρευνες στα μέσα κοινωνικής δικτύωσης για να παρέχουν πολύτιμες πληροφορίες για τις επιχειρήσεις και τους πελάτες.

Επισημάνση δεδομένων: Η βαθιά μάθηση απαιτεί δεδομένα με ετικέτα για την εκπαίδευση. Αφού εκπαιδευτεί, μπορεί να επισημάνει νέα δεδομένα και να προσδιορίσει διαφορετικούς τύπους δεδομένων από μόνο του.

Μηχανική χαρακτηριστικών: Ένας αλγόριθμος βαθιάς μάθησης μπορεί να εξοικονομήσει χρόνο επειδή δεν απαιτεί από τον άνθρωπο να εξάγει χαρακτηριστικά με μη αυτόματο τρόπο από ακατέργαστα δεδομένα.

Αποδοτικότητα: Όταν ένας αλγόριθμος βαθιάς μάθησης εκπαιδεύεται σωστά, μπορεί να εκτελέσει χιλιάδες εργασίες ξανά και ξανά, πιο γρήγορα από τους ανθρώπους.

Εκπαίδευση: Τα νευρωνικά δίκτυα που χρησιμοποιούνται στη βαθιά μάθηση έχουν την ικανότητα να εφαρμόζονται σε πολλούς διαφορετικούς τύπους δεδομένων και εφαρμογές. Επιπλέον, ένα μοντέλο βαθιάς μάθησης μπορεί να προσαρμοστεί επανεκπαιδευοντάς το με νέα δεδομένα.

ΚΕΦΑΛΑΙΟ 2^ο

Πώς λειτουργεί To deep learning ;

Τα προγράμματα υπολογιστών που χρησιμοποιούν deep learning περνούν σχεδόν από την ίδια διαδικασία με το μικρό παιδί που μαθαίνει να αναγνωρίζει τον σκύλο. Κάθε αλγόριθμος εφαρμόζει έναν μη γραμμικό μετασχηματισμό στην είσοδο του και δημιουργεί ένα στατιστικό μοντέλο ως έξοδο. Οι επαναλήψεις συνεχίζονται μέχρι η έξοδος φτάσει σε ένα καλό επίπεδο ακρίβειας.

Νευρωνικά δίκτυα και βαθιά μάθηση

Νευρωνικά δίκτυα, ένα όμορφο βιολογικά εμπνευσμένο πρότυπο προγραμματισμού που επιτρέπει σε έναν υπολογιστή να μαθαίνει από δεδομένα παρατήρησης. Τα νευρωνικά δίκτυα και η βαθιά μάθηση παρέχουν τις καλύτερες λύσεις σε πολλά προβλήματα που υπάρχουν στην αναγνώριση εικόνων, την αναγνώριση ομιλίας και την επεξεργασία φυσικής γλώσσας. Τα νευρωνικά δίκτυα είναι κόμβοι, όπως ακριβώς ο εγκέφαλος του ανθρώπου που αποτελείται από νευρώνες. Οι κόμβοι συνδέονται με γειτονικούς κόμβους. Σύμφωνα με τον αριθμό των επιπέδων του δικτύου, το δίκτυο θεωρείται βαθύτερο. Ένας νευρώνας στον εγκέφαλο του ανθρώπου λαμβάνει χιλιάδες σήματα από άλλους νευρώνες. Σε ένα τεχνητό νευρωνικό δίκτυο, τα σήματα μεταφέρονται μεταξύ των κόμβων και αποδίδονται τα αντίστοιχα βάρη. Ένας βαρύτερος κόμβος επηρεάζει το επόμενο στρώμα κόμβων. Το τελικό στρώμα συγκεντρώνει τις εισροές για να δημιουργήσει μια έξοδο. Τα συστήματα βαθιάς μάθησης απαιτούν αρκετό υλικό, επειδή έχουν πολλά δεδομένα για επεξεργασία και έχουν αρκετούς δύσκολους μαθηματικούς υπολογισμούς. Ωστόσο, ακόμη και με τόσο προηγμένο υλικό, η εκπαίδευση ενός νευρωνικού δικτύου μπορεί να διαρκέσει πολύ καιρό. Στην παραδοσιακή μηχανική μάθηση, ο προγραμματιστής πρέπει να είναι αναλυτικός όταν λέει στον υπολογιστή τι πράγματα πρέπει να αναζητήσει για να αποφασίσει εάν μια εικόνα περιέχει σκύλο ή όχι. Αυτή είναι μια δύσκολη και χρονοβόρα διαδικασία που ονομάζεται εξαγωγή χαρακτηριστικών και η επιτυχία εξαρτάται από την ικανότητα του προγραμματιστή να ορίζει με ακρίβεια ένα σύνολο χαρακτηριστικών για σκύλους. Το πλεονέκτημα της βαθιάς μάθησης είναι ότι το πρόγραμμα δίνει τη δυνατότητα που έχει οριστεί από μόνο του χωρίς επίβλεψη. Η μάθηση χωρίς επίβλεψη είναι ταχύτερη, αλλά και πιο ακριβής. Αρχικά, το πρόγραμμα

μπορεί να έχει δεδομένα εκπαίδευσης, δηλαδή ένα σύνολο από εικόνες για τις οποίες ένας άνθρωπος έχει επισημάνει αν σε μια εικόνα υπάρχει σκύλος ή όχι . Το πρόγραμμα χρησιμοποιεί πληροφορίες που του παρέχουν τα δεδομένα από την εκπαίδευση και δημιουργεί χαρακτηριστικά για σκύλους και δημιουργεί ένα μοντέλο πρόβλεψης.

Συγκεκριμένα, το μοντέλο που παράγει ο υπολογιστής μπορεί να προσδιορίσει οτιδήποτε στην εικόνα που έχει τέσσερα πόδια και μια ουρά θα πρέπει να ονομάζεται σκύλος. Φυσικά, αυτό δεν το γνωρίζει το πρόγραμμα . Απλώς θα αναζητήσει μοτίβα pixel στα ψηφιακά δεδομένα. Με κάθε επανάληψη, το μοντέλο πρόβλεψης γίνεται πιο περίπλοκο και πιο ακριβές. Σε αντίθεση με το νήπιο, που θα χρειαστεί εβδομάδες ή και μήνες για να κατανοήσει την έννοια του σκύλου, ένα πρόγραμμα υπολογιστή που χρησιμοποιεί αλγόριθμους βαθιάς μάθησης μπορεί να δείξει ένα σετ εκπαίδευσης και να ταξινομήσει εκατομμύρια εικόνες, προσδιορίζοντας με ακρίβεια ποιες εικόνες έχουν σκύλους μέσα σε λίγα λεπτά. Για να επιτευχθούν αποδεκτά επίπεδα με ακρίβεια, τα προγράμματα βαθιάς εκμάθησης απαιτούν πρόσβαση σε μεγάλες ποσότητες δεδομένων εκπαίδευσης και επεξεργαστικής ισχύος, την οποία οι προγραμματιστές δεν μπορούσαν εύκολα να αποκτήσουν πριν από την εποχή του cloud computing. Επειδή ο προγραμματισμός βαθιάς μάθησης μπορεί να δημιουργήσει πολύπλοκα στατιστικά μοντέλα απευθείας από τη δική του επαναληπτική έξοδο, είναι ικανός να παράγει ακριβή μοντέλα πρόβλεψης από μεγάλες ποσότητες μη επισημασμένων, μη δομημένων δεδομένων. Αυτό είναι σημαντικό καθώς το Διαδίκτυο των Πραγμάτων (IoT) γίνεται πιο διάχυτο, καθώς πολλά από τα δεδομένα που παράγονται από ανθρώπους και μηχανές είναι αδόμητα και χωρίς ετικέτα. Τα νευρωνικά δίκτυα βαθιάς μάθησης μιμούνται τον ανθρώπινο εγκέφαλο μέσω μιας εισροής δεδομένων, βαρών και προκαταλήψεων. Αυτά τα στοιχεία συμβάλλουν στην συνεργασία για την αναγνώριση, ταξινόμηση και περιγραφή αντικειμένων στα δεδομένα σας. Το νευρωνικό δίκτυο αποτελείται από πολλά επίπεδα συνδεδεμένων κόμβων, με κάθε κόμβο να έχει εξάρτηση από το προηγούμενο επίπεδο για να βελτιωθεί και τη βελτιστοποιηθεί η πρόβλεψη ή η ταξινόμηση.

Η υπολογιστική διαδικασία μέσω αυτού του νευρωνικού δικτύου ονομάζεται μπροστινή διάδοση. Η είσοδος και η έξοδος ενός νευρωνικού δικτύου ονομάζονται ορατά στρώματα. Η είσοδος είναι εκεί που το μοντέλο βαθιάς μάθησης λαμβάνει δεδομένα τα οποία επεξεργάζεται, και η έξοδος είναι εκεί που γίνονται οι τελικές

προβλέψεις. Τα παραπάνω περιγράφουν έναν απλό τύπο νευρωνικού δικτύου με απλούς όρους. Οι αλγόριθμοι βαθιάς μάθησης είναι πολύ περίπλοκοι και υπάρχουν αρκετοί τύποι Νευρωνικών Δικτύων για την λύση συγκεκριμένων προβλημάτων ή συνόλων δεδομένων.

Για παράδειγμα, τα συνελκτικά νευρωνικά δίκτυα, χρησιμοποιούνται σε εφαρμογές υπολογιστών και ταξινόμησης σε εικόνες, μπορούν επίσης να ανιχνεύσουν χαρακτηριστικά μέσα σε μια εικόνα και επιτρέπουν εργασίες, όπως το να ανιχνεύουν ή να αναγνωρίζουν αντικείμενα. Τη χρονιά 2015, ένα συνελκτικό νευρωνικό δίκτυο σε μια πρόκληση αναγνώρισης αντικειμένων για πρώτη φορά ξεπέρασε έναν άνθρωπο. Τα επαναλαμβανόμενα νευρωνικά δίκτυα χρησιμοποιούνται σε εφαρμογές φυσικής γλώσσας όπως και στην αναγνώριση ομιλίας, αφού αξιοποιεί τα δεδομένα διαδοχικών ή χρονικών σειρών. Οι περισσότερες μέθοδοι βαθιάς μάθησης έχουν χρησιμοποιήσει πολλές αρχιτεκτονικές νευρωνικών δικτύων, αλλά και τα μοντέλα βαθιάς μάθησης χαρακτηρίζονται ως βαθιά νευρωνικά δίκτυα. Ο όρος "βαθύς" αναφέρεται σε έναν αριθμό από κρυφών επιπέδων σε ένα νευρωνικό δίκτυο. Τα παραδοσιακά νευρωνικά δίκτυα περιέχουν ελάχιστα κρυφά επίπεδα, ενώ τα βαθιά δίκτυα μπορούν να έχουν πάνω από 100. Τα μοντέλα βαθιάς μάθησης έχουν εκπαιδευτεί χρησιμοποιώντας μεγάλο αριθμό από σύνολα δεδομένων με αρχιτεκτονικές νευρωνικών δικτύων που μαθαίνουν τα χαρακτηριστικά κατευθείαν από τα δεδομένα χωρίς να χρησιμοποιούν χειροκίνητα την εξαγωγή χαρακτηριστικών.

Γιατί Βαθιά Μάθηση;

Με μια λέξη, Ακρίβεια. Τα μοντέλα βαθιάς μάθησης βελτιώνουν την ακρίβειά τους καθώς μεγαλώνει ο αριθμός των δεδομένων εκπαίδευσης. Βασικό πλεονέκτημα αυτών είναι ότι βελτιώνονται με το πέρασμα του χρόνου καθώς αυξάνεται ο αριθμός των δεδομένων. Οι πιο πρόσφατες εξελίξεις στη βαθιά μάθηση έχουν βελτιωθεί τόσο, που η βαθιά μάθηση έχει ξεπεράσει τους ανθρώπους σε ορισμένες εργασίες, όπως στην ταξινόμηση αντικειμένων σε εικόνες.

ΚΕΦΑΛΑΙΟ 3^ο

Άνοδος της βαθιάς μάθησης

Λέγεται ότι η μηχανική μάθηση συνέβη τη δεκαετία του 1950 όταν ο Άλαν Τούρινγκ, ένας Βρετανός μαθηματικός, πρότεινε την τεχνητά ευφυή «μηχανή μάθησης» του. Ο Άρθουρ Σάμουελ έγραψε το πρώτο πρόγραμμα εκμάθησης υπολογιστών. Το πρόγραμμά του έκανε έναν υπολογιστή της IBM να βελτιώνεται στο παιχνίδι πούλια όσο περισσότερο έπαιζε. Τις δεκαετίες που ακολούθησαν, διάφορες τεχνικές μηχανικής μάθησης μπήκαν και βγήκαν από τη μόδα. Τα νευρωνικά δίκτυα αγνοήθηκαν ως επί το πλείστον από τους ερευνητές μηχανικής μάθησης, καθώς μαστίζονταν από το πρόβλημα των «τοπικών ελάχιστων» στο οποίο οι σταθμίσεις εσφαλμένα φαινόταν να δίνουν τα λιγότερα σφάλματα. Ωστόσο, ορισμένες τεχνικές μηχανικής εκμάθησης όπως η όραση υπολογιστή και η αναγνώριση προσώπου προχώρησαν. Το 2001, αναπτύχθηκε ένας αλγόριθμος μηχανικής μάθησης που ονομάζεται Adaboost για να ανιχνεύει πρόσωπα μέσα σε μια εικόνα σε πραγματικό χρόνο. Φιλτράρει τις εικόνες μέσω συνόλων αποφάσεων όπως "έχει η εικόνα ένα φωτεινό σημείο ανάμεσα σε σκοτεινά μπαλώματα, που πιθανώς υποδηλώνει τη γέφυρα μιας μύτης;" Όταν τα δεδομένα μετακινήθηκαν πιο κάτω στο δέντρο αποφάσεων, η πιθανότητα επιλογής του σωστού προσώπου από μια εικόνα μεγάλωνε.

Τα νευρωνικά δίκτυα δεν επέστρεψαν στην εύνοια για αρκετά ακόμη χρόνια, όταν τελικά μπήκαν στην αγορά ισχυρές μονάδες επεξεργασίας γραφικών. Οι νέοι ερευνητές με δυνατότητα υλικού χρησιμοποιούν επιτραπέζιους υπολογιστές για την εκτέλεση, τον χειρισμό και την επεξεργασία εικόνων. Το πιο σημαντικό άλμα προς τα εμπρός για τα νευρωνικά δίκτυα συνέβη λόγω της εισαγωγής σημαντικών ποσοτήτων επισημασμένων δεδομένων με το ImageNet, μια βάση δεδομένων με εκατομμύρια ετικέτες εικόνων από το Διαδίκτυο. Η δυσκίνητη εργασία της χειροκίνητης επισήμανσης εικόνων αντικαταστάθηκε από το crowdsourcing, δίνοντας στα δίκτυα μια ουσιαστικά απεριόριστη πηγή εκπαιδευτικού υλικού. Στα χρόνια από τότε που οι εταιρείες τεχνολογίας έκαναν τις βιβλιοθήκες βαθιάς μάθησης ανοιχτού κώδικα. Παραδείγματα περιλαμβάνουν το Google Tensorflow, τις λειτουργικές μονάδες ανοιχτού κώδικα του Facebook για το Torch, το Amazon DSSTNE στο GitHub και το Microsoft CNTK.

Δράση του Deep Learning

Εκτός από την αγαπημένη σας υπηρεσία ροής μουσικής που προτείνει μελωδίες που μπορείτε να απολαύσετε, πώς επηρεάζει η βαθιά μάθηση τις ζωές των ανθρώπων; Όπως αποδεικνύεται, η βαθιά μάθηση βρίσκει το δρόμο της σε εφαρμογές όλων των μεγεθών. Οποιοσδήποτε χρησιμοποιεί το Facebook δεν μπορεί παρά να παρατηρήσει ότι η πλατφόρμα κοινωνικής δικτύωσης συνήθως αναγνωρίζει και επισημαίνει τους φίλους σας όταν ανεβάζετε νέες φωτογραφίες. Οι ψηφιακοί βοηθοί όπως οι Siri, Cortana, Alexa και Google Now χρησιμοποιούν βαθιά εκμάθηση για επεξεργασία φυσικής γλώσσας και αναγνώριση ομιλίας. Το Skype μεταφράζει προφορικές συνομιλίες σε πραγματικό χρόνο. Πολλές πλατφόρμες ηλεκτρονικού ταχυδρομείου έχουν καταστεί ικανές να αναγνωρίζουν τα ανεπιθύμητα μηνύματα πριν καν φτάσουν στα εισερχόμενα. Το PayPal έχει εφαρμόσει βαθιά εκμάθηση για να αποτρέψει τις δόλιες πληρωμές. Εφαρμογές όπως το CamFind επιτρέπουν στους χρήστες να τραβήξουν μια φωτογραφία οποιουδήποτε αντικειμένου και, χρησιμοποιώντας την τεχνολογία οπτικής αναζήτησης για φορητές συσκευές, να ανακαλύψουν ποιο είναι το αντικείμενο.

Η Google, ειδικότερα, αξιοποιεί τη βαθιά μάθηση για να προσφέρει λύσεις. Το πρόγραμμα υπολογιστή AlphaGo της Google Deepmind νίκησε πρόσφατα τους μόνιμους πρωταθλητές στο παιχνίδι Go. Το WaveNet της DeepMind μπορεί να δημιουργήσει ομιλία που μιμείται την ανθρώπινη φωνή που ακούγεται πιο φυσικό από τα συστήματα ομιλίας που υπάρχουν σήμερα στην αγορά. Το Google Translate χρησιμοποιεί τη βαθιά εκμάθηση και την αναγνώριση εικόνας για τη μετάφραση φωνητικών και γραπτών γλωσσών. Το Google Planet μπορεί να αναγνωρίσει πού τραβήχτηκε οποιαδήποτε φωτογραφία. Η Google ανέπτυξε τη βάση δεδομένων λογισμικού βαθιάς μάθησης, Tensorflow, για να βοηθήσει στην παραγωγή εφαρμογών AI. Τα αυτό-οδηγούμενα αυτοκίνητα δοκιμάζονται παγκοσμίως. το πολύπλοκο στρώμα των νευρωνικών δικτύων εκπαιδεύεται για να προσδιορίζει αντικείμενα προς αποφυγή, να αναγνωρίζει τα φανάρια και να γνωρίζει πότε να προσαρμόζει την ταχύτητα. Τα νευρωνικά δίκτυα γίνονται ικανά να προβλέπουν τα πάντα, από τις τιμές των μετοχών μέχρι τον καιρό.

Σκεφτείτε την αξία των ψηφιακών βοηθών που μπορούν να προτείνουν πότε να πουλήσετε μετοχές ή πότε να εκκενώσετε πριν από έναν τυφώνα. Οι εφαρμογές βαθιάς

μάθησης θα σώσουν ακόμη και ζωές καθώς αναπτύσσουν την ικανότητα να σχεδιάζουν σχέδια θεραπείας βασισμένων σε στοιχεία για ιατρικούς ασθενείς και να βοηθούν στην έγκαιρη ανίχνευση του καρκίνου.

Παραδείγματα βαθιάς μάθησης

Επειδή τα μοντέλα βαθιάς μάθησης επεξεργάζονται πληροφορίες με τρόπους παρόμοιους με τον ανθρώπινο εγκέφαλο, μπορούν να εφαρμοστούν σε πολλές εργασίες που κάνουν οι άνθρωποι. Η βαθιά εκμάθηση χρησιμοποιείται επί του παρόντος στα περισσότερα εργαλεία που αναγνωρίζουν εικόνες, στην επεξεργασία της φυσικής γλώσσας όπως και στην αναγνώριση ομιλίας. Τα εργαλεία αυτά έχουν αρχίσει να εμφανίζονται σε πολλές εφαρμογές που είναι πολύ διαφορετικές μεταξύ τους, όπως είναι τα αυτόνομα αυτοκίνητα και οι υπηρεσίες μετάφρασης γλωσσών. Οι περιπτώσεις χρήσης σήμερα για βαθιά μάθηση περιλαμβάνουν όλους τους τύπους εφαρμογών ανάλυσης μεγάλων δεδομένων, ειδικά εκείνες που επικεντρώνονται στο NLP, τη μετάφραση γλώσσας, την ιατρική διάγνωση, τα σήματα συναλλαγών στο χρηματιστήριο, την ασφάλεια δικτύου και την αναγνώριση εικόνας. Οι εφαρμογές της βαθιάς μάθησης έχουν χρησιμοποιηθεί σε βιομηχανίες τόσο σε Αυτόνομη οδήγηση όσο και σε ιατρικές συσκευές.

Αυτόνομη οδήγηση: Οι ερευνητές αυτοκινήτων χρησιμοποιούν βαθιά εκμάθηση για να ανιχνεύουν αυτόματα αντικείμενα όπως πινακίδες στάσης και φανάρια. Επιπλέον, η βαθιά μάθηση χρησιμοποιείται για τον εντοπισμό πεζών, συμβάλλοντας στη μείωση των ατυχημάτων.

Διαστημική και Άμυνα: Η βαθιά εκμάθηση χρησιμοποιείται για την ανίχνευση δορυφορικών αντικειμένων που εντοπίζουν περιοχές ενδιαφέροντος και εντοπίζουν ασφαλείς ή μη ασφαλείς περιοχές για τα στρατεύματα.

Ιατρική έρευνα: Οι ερευνητές του καρκίνου χρησιμοποιούν τη βαθιά μάθηση για να ανιχνεύουν τα καρκινικά κύτταρα. Η ομάδα του UCLA κατασκεύασε ένα μικροσκόπιο που μπορεί να αποδώσει ένα σύνολο δεδομένων με υψηλές διαστάσεις για την εκπαίδευση μιας εφαρμογής βαθιάς μάθησης για την ακριβή αναγνώριση των καρκινικών κυττάρων.

Βιομηχανικός αυτοματισμός: Η βαθιά μάθηση έχει βελτιώσει την ασφάλεια των εργαζομένων γύρω από επικίνδυνα μηχανήματα, εντοπίζοντας αυτόματα πότε άτομα ή αντικείμενα βρίσκονται σε μη ασφαλή απόσταση από μηχανήματα.

Ηλεκτρονικά: Η βαθιά μάθηση έχει χρησιμοποιηθεί στην αυτόνομη μετάφραση ακοής και ομιλίας. Για παράδειγμα, μια οικιακή βοηθητική συσκευή που ανταποκρίνεται στη φωνή σας και μαθαίνει τις προτιμήσεις σας τροφοδοτείται από εφαρμογές βαθιάς εκμάθησης. Συγκεκριμένα πεδία στα οποία χρησιμοποιείται επί του παρόντος η βαθιά μάθηση περιλαμβάνουν τα ακόλουθα:

Εμπειρία πελάτη (CX). Τα μοντέλα Deep Learning χρησιμοποιούνται ήδη για chatbots. Και, καθώς συνεχίζει να ωριμάζει, η βαθιά μάθηση αναμένεται να εφαρμοστεί σε διάφορες επιχειρήσεις ώστε να αυξήσουν της ικανοποίησης των πελατών.

Δημιουργία κειμένου. Οι μηχανές διδάσκονται τη γραμματική και το στυλ ενός κειμένου και στη συνέχεια χρησιμοποιούν αυτό το μοντέλο για να δημιουργήσουν αυτόματα ένα εντελώς νέο κείμενο που ταιριάζει με τη σωστή ορθογραφία, γραμματική και στυλ του αρχικού κειμένου.

Προσθήκη χρώματος. Το χρώμα μπορεί να προστεθεί σε ασπρόμαυρες φωτογραφίες και βίντεο χρησιμοποιώντας μοντέλα βαθιάς εκμάθησης. Στο παρελθόν, αυτή ήταν μια εξαιρετικά χρονοβόρα, χειρωνακτική διαδικασία.

Υπολογιστική όραση. Η βαθιά εκμάθηση έχει βελτιώσει σε μεγάλο βαθμό τους υπολογιστές, παρέχοντας στους υπολογιστές ακρίβεια για τον εντοπισμό αντικειμένων και επίσης ταξινομούν, αποκαθιστούν και τμηματοποιούν τις εικόνες.

ΚΕΦΑΛΑΙΟ 5^ο

Βαθιά μάθηση έναντι μηχανικής μάθησης

Η βαθιά μάθηση είναι ένα σύνολο της μηχανικής μάθησης που ξεχωρίζει από την προσέγγισή της στην επίλυση προβλημάτων. Η μηχανική εκμάθηση απαιτεί από τους ειδικούς του τομέα να προσδιορίσουν το μεγαλύτερο μέρος της λειτουργικότητας της εφαρμογής. Αντίθετα, η βαθιά μάθηση καταργεί την ανάγκη για εξειδίκευση στον τομέα. Αυτό επιτρέπει στη βαθιά μάθηση να χρειάζεται πολύ περισσότερο χρόνο για να εκπαιδευτεί από τη μηχανική μάθηση, οι οποίοι χρειάζονται δευτερόλεπτα έως ώρες. Ωστόσο, το αντίθετο ίσχυε κατά τη διάρκεια των δοκιμών. Οι αλγόριθμοι βαθιάς μάθησης χρειάζονται λιγότερο χρόνο για την να εκτελέσουν δοκιμές από τους αλγόριθμους μηχανικής μάθησης και ο χρόνος δοκιμής για τους αλγόριθμους μηχανικής μάθησης αυξάνεται με το μέγεθος των δεδομένων. Επιπλέον, η μηχανική εκμάθηση δεν απαιτεί τις ίδιες δαπανηρές, προηγμένες μηχανές και GPU υψηλής απόδοσης που απαιτεί η βαθιά εκμάθηση.

Στο τέλος, πολλοί επιστήμονες δεδομένων επιλέγουν την παραδοσιακή μηχανική μάθηση έναντι της βαθιάς μάθησης λόγω της ανώτερης ερμηνευτικότητάς της ή της ικανότητας να κατανοούν τις λύσεις. Οι αλγόριθμοι μηχανικής εκμάθησης προτιμώνται επίσης όταν τα δεδομένα είναι μικρά. Οι περιπτώσεις όπου η βαθιά μάθηση γίνεται προτιμητέα περιλαμβάνουν καταστάσεις όπου υπάρχει μεγάλος όγκος δεδομένων, έλλειψη κατανόησης του τομέα για ενδοσκόπηση χαρακτηριστικών ή πολύπλοκα προβλήματα, όπως η αναγνώριση ομιλίας και το NLP.

Η βαθιά μάθηση είναι μια μορφή μηχανικής μάθησης. Μια ροή εργασιών μηχανικής εκμάθησης ξεκινά με τη μη αυτόματη εξαγωγή σχετικών λειτουργιών από μια εικόνα. Αυτά τα χαρακτηριστικά χρησιμοποιούνται στη συνέχεια για την κατασκευή ενός μοντέλου που ταξινομεί αντικείμενα σε εικόνες. Οι σχετικές λειτουργίες μπορούν να εξαχθούν αυτόματα από τις εικόνες μέσω μιας ροής εργασιών βαθιάς εκμάθησης. Επιπλέον, η βαθιά μάθηση εκτελεί «από άκρο σε άκρο μάθηση». Στο δίκτυο δίνονται ακατέργαστα δεδομένα και μια εργασία προς εκτέλεση (όπως ταξινόμηση) και μαθαίνει αυτόματα πώς να το κάνει.

Μια διαδικασία που ονομάζεται *backpropagation* χρησιμοποιεί συγκεκριμένους αλγόριθμους όπως η *gradient descent* για να υπολογίσει το σφάλμα και προσαρμόζει τα βάρη και τις προκαταλήψεις της συνάρτησης μετακινώντας τα στρώματα προς τα πίσω σε μια προσπάθεια εκπαίδευσης του μοντέλου. Μαζί, η προς τα εμπρός διάδοση και η αντίστροφη διάδοση επιτρέπουν στο νευρωνικό δίκτυο να κάνει προβλέψεις και να διορθώνει ανάλογα τυχόν σφάλματα. Ένα βασικό πλεονέκτημα των δικτύων βαθιάς μάθησης είναι ότι συνήθως βελτιώνονται καθώς αυξάνονται τα δεδομένα.



Μέθοδοι Βαθιάς Μάθησης

Μια ποικιλία προσεγγίσεων μπορεί να χρησιμοποιηθεί για τη δημιουργία ισχυρών μοντέλων βαθιάς μάθησης, συμπεριλαμβανομένων τεχνικών όπως η μείωση του ποσοστού μάθησης, η μεταφορά μάθησης, η εκπαίδευση από την αρχή και η εγκατάλειψη.

Μείωση του ρυθμού της μάθησης: Ο ρυθμός εκμάθησης είναι μια υπερπαράμετρος - ένας παράγοντας που καθορίζει ένα σύστημα ή θέτει τις συνθήκες κάτω από τις οποίες λειτουργεί πριν από τη διαδικασία εκμάθησης - που ελέγχει πόσο ανταποκρίνεται το μοντέλο στο σφάλμα εκτίμησης κάθε φορά που αλλάζουν τα βάρη του μοντέλου. Πολύ υψηλός ρυθμός μάθησης μπορεί να οδηγήσει σε μια ασταθή προπονητική διαδικασία ή να μάθει ένα μη βέλτιστο σύνολο βαρών. Τα ποσοστά μάθησης που είναι πολύ υψηλά μπορεί να οδηγήσουν σε ασταθείς προπονητικές διαδικασίες ή στην εκμάθηση ενός μη βέλτιστου συνόλου βαρών. Ένας πολύ μικρός ρυθμός μάθησης θα οδηγήσει σε μια πολύ μακρά διαδικασία εκπαίδευσης, η οποία μπορεί να κολλήσει. Οι μέθοδοι μείωσης του ρυθμού μάθησης (γνωστές επίσης ως ανόπτηση ρυθμού μάθησης ή προσαρμοστικός ρυθμός μάθησης) είναι η διαδικασία προσαρμογής του ρυθμού μάθησης για τη βελτίωση της απόδοσης και τη μείωση του χρόνου εκπαίδευσης. Οι πιο εύκολες και πιο συνηθισμένες προσαρμογές του ρυθμού μάθησης κατά τη διάρκεια της

εκπαίδευσης περιλαμβάνουν τεχνικές για τη μείωση του ρυθμού μάθησης με την πάροδο του χρόνου.

Εκμάθηση της μεταφοράς: Αυτή η διαδικασία περιλαμβάνει τη βελτίωση ενός προηγούμενως εκπαιδευμένου μοντέλου. Απαιτεί μια διεπαφή στο εσωτερικό ενός προϋπάρχοντος δικτύου. Πρώτον, οι χρήστες παρέχουν στο υπάρχον δίκτυο νέα δεδομένα που περιέχουν προηγούμενως άγνωστες ταξινομήσεις. Μόλις συντονιστεί το δίκτυο, μπορεί να εκτελέσει νέες εργασίες με πιο συγκεκριμένες δυνατότητες ταξινόμησης. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι απαιτεί πολύ λιγότερα δεδομένα από άλλες μεθόδους, μειώνοντας το χρόνο υπολογισμού σε λεπτά ή ώρες.

Εξάσκηση από την αρχή: Αυτή η προσέγγιση απαιτεί από τους προγραμματιστές να συλλέγουν μεγάλες ποσότητες δεδομένων με ετικέτα και να δημιουργούν αρχιτεκτονικές δικτύου που μπορούν να μάθουν χαρακτηριστικά και μοντέλα. Αυτή η τεχνική είναι ιδιαίτερα χρήσιμη για νέες εφαρμογές και για εκείνες με μεγάλο αριθμό κατηγοριών εξόδου. Σε γενικές γραμμές, ωστόσο, αυτή είναι μια λιγότερο συνηθισμένη προσέγγιση γιατί απαιτεί πολλά δεδομένα, με αποτέλεσμα ημέρες ή εβδομάδες για προπόνηση.

Εγκατάλειψη: Αυτή η μέθοδος επιχειρεί να λύσει το πρόβλημα της υπερπροσαρμογής σε δίκτυα με μεγάλο αριθμό παραμέτρων με τυχαία απόρριψη μονάδων και των συνδέσεών τους από το νευρωνικό δίκτυο κατά τη διάρκεια της εκπαίδευσης. Οι μέθοδοι εγκατάλειψης έχουν αποδειχθεί ότι βελτιώνουν την απόδοση των νευρωνικών δικτύων σε εποπτευόμενες μαθησιακές εργασίες σε τομείς όπως η αναγνώριση ομιλίας, η ταξινόμηση εγγράφων και η υπολογιστική βιολογία.

ΚΕΦΑΛΑΙΟ 6^ο

Προοπτικές σταδιοδρομίας Deep Learning

Ο τομέας της τεχνητής νοημοσύνης είναι σοβαρά υποστελεχωμένος. Αν και δεν προσλαμβάνουν ακόμη όλες οι εταιρείες επαγγελματίες με δεξιότητες βαθιάς μάθησης, η ύπαρξη τέτοιων εκπαιδευμένων ειδικών αναμένεται να γίνει σταδιακά μια κρίσιμη απαίτηση για οργανισμούς που θέλουν να παραμείνουν ανταγωνιστικοί και να προωθήσουν την καινοτομία. Οι μηχανικοί μηχανικής μάθησης έχουν μεγάλη ζήτηση επειδή ούτε οι επιστήμονες δεδομένων ούτε οι μηχανικοί λογισμικού έχουν ακριβώς τις δεξιότητες που απαιτούνται για τον τομέα της μηχανικής μάθησης. Ο ρόλος του μηχανικού μηχανικής μάθησης έχει εξελιχθεί για να καλύψει το κενό. Τι πολλά υποσχόμενη είναι η βαθιά μάθηση όσον αφορά τις ευκαιρίες σταδιοδρομίας και τις αμοιβές; Αρκετά. Το Glassdoor αναφέρει τον μέσο μισθό για έναν μηχανικό μηχανικής εκμάθησης σε σχεδόν 115.000 \$ ετησίως. Σύμφωνα με το PayScale, το εύρος μισθών κυμαίνεται από 100.000 \$ έως 166.000 \$. Η ανάπτυξη θα επιταχυνθεί τα επόμενα χρόνια καθώς τα συστήματα και τα εργαλεία βαθιάς μάθησης βελτιώνονται και επεκτείνονται σε όλους τους κλάδους.

Πλεονεκτήματα Της Βαθιάς Μάθησης

Το καλύτερο με τη βαθιά μάθηση είναι ότι μπορεί να εφαρμοστεί σε τόσο μεγάλο εύρος περιοχών, πράγμα που σημαίνει ότι υπάρχουν πολλές διαφορετικές εφαρμογές για αυτόν τον τύπο τεχνητής νοημοσύνης. Αυτοί οι τύποι δικτύων λαμβάνουν τόσες πολλές πληροφορίες και τις αποσπάζουν σε κατανοητά κομμάτια, ενώ συγκρίνουν επίσης νέα δεδομένα με όσα έχουν ήδη μάθει στο παρελθόν, γεγονός που τους επιτρέπει να επιτύχουν απίστευτα ακριβή αποτελέσματα με πολύ λίγη ανθρώπινη παρέμβαση. Το μόνο μειονέκτημα αυτή τη στιγμή είναι ότι αυτοί οι αλγόριθμοι συχνά αδυνατούν ή δεν θέλουν να εξηγήσουν γιατί κάτι λειτούργησε όπως έγινε, επειδή ακόμα δεν γνωρίζουμε αρκετά για το πώς λειτουργεί ο εγκέφαλος τους, αλλά ελπίζουμε, καθώς η έρευνα συνεχίζεται. Θα φτάσουμε τελικά σε ένα σημείο όπου ακόμη και τα πιο προηγμένα μοντέλα μηχανικής μάθησης γίνονται αρκετά έξυπνα ώστε να μας πουν ακριβώς τι περνάει από το μυαλό τους κάθε δεδομένη στιγμή.

Αλγόριθμοι βαθιάς μάθησης

Οι αλγόριθμοι βαθιάς μάθησης εξάγουν θεωρίες όπως ο ανθρώπινος εγκέφαλος, και αναλύουν συστηματικά τα δεδομένα με μια λογική δομή. Για να ολοκληρωθεί αυτό, χρησιμοποιούν μια πολυεπίπεδη δομή αλγορίθμων που ονομάζονται Νευρωνικά Δίκτυα.

Η σχεδίαση του νευρωνικού δικτύου είναι βασισμένη στη δομή του ανθρώπινου εγκέφαλου. Όπως εμείς χρησιμοποιούμε το μυαλό για να εντοπίσουμε χαρακτηριστικά και ταξινομούμε διαφορετικούς τύπους πληροφοριών, έτσι και τα νευρωνικά δίκτυα διδάσκονται και εκτελούν τις ίδιες εργασίες στα δεδομένα. Ο εγκέφαλος του ανθρώπου λειτουργεί σχεδόν το ίδιο. Κάθε φορά που λαμβάνει νέες πληροφορίες, ο εγκέφαλος προσπαθεί να τις παρομοιάσει και να τις συγκρίνει με άλλα γνωστά αντικείμενα.

Εμπόδια στην εφαρμογή της βαθιάς μάθησης

Ενώ αποκαλύπτονται νέες χρήσεις για τη βαθιά μάθηση, εξακολουθεί να είναι ένα εξελισσόμενο πεδίο με ορισμένους περιορισμούς:

Μεγάλοι όγκοι δεδομένων: Προκειμένου να επιτευχθούν πιο διορατικές και αφηρημένες απαντήσεις, η βαθιά μάθηση απαιτεί μεγάλες ποσότητες δεδομένων για εκπαίδευση. Παρόμοια με τον ανθρώπινο εγκέφαλο, ένας αλγόριθμος βαθιάς μάθησης χρειάζεται παραδείγματα ώστε να μπορεί βελτιώσει το αποτέλεσμά του μέσα από τα λάθη.

Έλλειψη ευελιξίας: Οι μηχανές εξακολουθούν να μαθαίνουν με πολύ στενούς τρόπους, κάτι που μπορεί να οδηγήσει σε λάθη. Τα δίκτυα βαθιάς μάθησης χρειάζονται δεδομένα για να λύσουν ένα συγκεκριμένο πρόβλημα. Εάν σας ζητηθεί να εκτελέσετε μια εργασία εκτός αυτού του πεδίου εφαρμογής, πιθανότατα θα αποτύχει.

Έλλειψη διαφάνειας: Ενώ κοσκινίζει εκατομμύρια σημεία δεδομένων για να βρει μοτίβα, μπορεί να είναι δύσκολο να κατανοήσουμε πώς ένα νευρωνικό δίκτυο φτάνει στη λύση του. Αυτή η έλλειψη διαφάνειας στον τρόπο με τον οποίο επεξεργάζονται δεδομένα καθιστά δύσκολο τον εντοπισμό ανεπιθύμητων προκαταλήψεων και την εξήγηση των προβλέψεων.

Παρά αυτά τα εμπόδια, οι επιστήμονες δεδομένων πλησιάζουν όλο και περισσότερο στη δημιουργία μοντέλων βαθιάς μάθησης υψηλής ακρίβειας που μπορούν να μάθουν χωρίς επίβλεψη – κάτι που θα κάνει τη βαθιά μάθηση ταχύτερη και λιγότερο εντατική.

ΚΕΦΑΛΑΙΟ 7^ο

Απαιτήσεις υλικού Deep Learning

Η βαθιά μάθηση απαιτεί τεράστια υπολογιστική ισχύ. Οι μονάδες γραφικής επεξεργασίας με υψηλή απόδοση είναι ιδανικές επειδή χειρίζονται μεγάλο αριθμό υπολογισμών σε πολλούς πυρήνες με άφθονη διαθέσιμη μνήμη. Ωστόσο, η διαχείριση πολλαπλών GPU επί τόπου μπορεί να δημιουργήσει μεγάλη ζήτηση σε εσωτερικούς πόρους και να είναι απίστευτα δαπανηρή σε κλίμακα.

Deep Learning και IBM

Για δεκαετίες τώρα, η IBM είναι πρωτοπόρος στην ανάπτυξη τεχνολογιών τεχνητής νοημοσύνης και βαθιάς μάθησης, που τονίζεται από την ανάπτυξη του IBM Watson, του chatbot AI της IBM. Ένα από τα πρώτα επιτεύγματα στην τεχνολογία βαθιάς μάθησης, η Watson είναι πλέον μια αξιόπιστη λύση για επιχειρήσεις που επιθυμούν να εφαρμόσουν προηγμένες τεχνικές επεξεργασίας φυσικής γλώσσας και μηχανικής μάθησης στα συστήματά τους χρησιμοποιώντας μια αποδεδειγμένη κλιμακωτή προσέγγιση για την υιοθέτηση και την εφαρμογή της τεχνητής νοημοσύνης.

Η Watson χρησιμοποιεί το πλαίσιο Apache Unstructured Information Management Architecture (UIMA) και το λογισμικό DeepQA της IBM για να καταστήσει διαθέσιμες στις εφαρμογές ισχυρές δυνατότητες βαθιάς εκμάθησης. Χρησιμοποιώντας εργαλεία όπως το IBM Watson Studio, η επιχείρησή σας μπορεί να αξιοποιήσει τα μεγάλα δεδομένα σας και να φέρει σε παραγωγή τα έργα επιστήμης δεδομένων σας, ενώ αναπτύσσει και εκτελεί τα μοντέλα σας σε οποιοδήποτε cloud.

Περιορισμοί και προκλήσεις

Ο μεγαλύτερος περιορισμός των μοντέλων βαθιάς μάθησης είναι ότι μαθαίνουν μέσω παρατηρήσεων. Αυτό σημαίνει ότι γνωρίζουν μόνο τι υπήρχε στα δεδομένα στα οποία εκπαιδεύτηκαν. Εάν ένας χρήστης έχει μικρό όγκο δεδομένων ή προέρχεται από μια συγκεκριμένη πηγή που δεν είναι απαραίτητα αντιπροσωπευτική της ευρύτερης λειτουργικής περιοχής, τα μοντέλα δεν θα μάθουν με τρόπο που να είναι γενικεύσιμος.

Το ζήτημα των προκαταλήψεων είναι επίσης ένα σημαντικό πρόβλημα για τα μοντέλα βαθιάς μάθησης. Εάν ένα μοντέλο εκπαιδεύεται σε δεδομένα που περιέχουν προκαταλήψεις, το μοντέλο θα αναπαράγει αυτές τις προκαταλήψεις στις προβλέψεις του. Αυτό ήταν ένα ενοχλητικό πρόβλημα για τους προγραμματιστές βαθιάς μάθησης, επειδή τα μοντέλα μαθαίνουν να διαφοροποιούνται με βάση τις ανεπαίσθητες παραλλαγές στα στοιχεία δεδομένων. Συχνά, οι παράγοντες που καθορίζει ότι είναι σημαντικοί δεν γίνονται ρητά σαφείς στον προγραμματιστή. Αυτό σημαίνει, για παράδειγμα, ένα μοντέλο αναγνώρισης προσώπου μπορεί να κάνει προσδιορισμούς για τα χαρακτηριστικά των ανθρώπων με βάση πράγματα όπως η φυλή ή το φύλο χωρίς να το γνωρίζει ο προγραμματιστής.

Το ποσοστό μάθησης μπορεί επίσης να γίνει μια σημαντική πρόκληση για τα μοντέλα βαθιάς μάθησης. Εάν το ποσοστό είναι πολύ υψηλό, τότε το μοντέλο θα συγκλίνει πολύ γρήγορα, παράγοντας μια λιγότερο από τη βέλτιστη λύση. Εάν το ποσοστό είναι πολύ χαμηλό, τότε η διαδικασία μπορεί να κολλήσει και θα είναι ακόμη πιο δύσκολο να βρεθεί λύση. Οι απαιτήσεις υλικού για τα μοντέλα βαθιάς εκμάθησης μπορούν επίσης να δημιουργήσουν περιορισμούς. Απαιτούνται πολυπύρηνες μονάδες επεξεργασίας γραφικών υψηλής απόδοσης (GPU) και άλλες παρόμοιες μονάδες επεξεργασίας για τη διασφάλιση βελτιωμένης απόδοσης και μειωμένης κατανάλωσης χρόνου. Ωστόσο, αυτές οι μονάδες είναι ακριβές και χρησιμοποιούν μεγάλες ποσότητες ενέργειας.

Άλλοι περιορισμοί και προκλήσεις περιλαμβάνουν τα ακόλουθα:

- Η βαθιά μάθηση απαιτεί μεγάλες ποσότητες δεδομένων. Επιπλέον, τα πιο ισχυρά και ακριβή μοντέλα θα χρειαστούν περισσότερες παραμέτρους, οι οποίες με τη σειρά τους απαιτούν περισσότερα δεδομένα.
- Μόλις εκπαιδευτούν, τα μοντέλα βαθιάς μάθησης γίνονται άκαμπτα και δεν μπορούν να χειριστούν το multitasking. Μπορούν να προσφέρουν αποτελεσματικές και ακριβείς λύσεις αλλά μόνο σε ένα συγκεκριμένο πρόβλημα. Ακόμη και η επίλυση ενός παρόμοιου προβλήματος θα απαιτούσε επανεκπαίδευση του συστήματος.

Οποιαδήποτε εφαρμογή απαιτεί συλλογισμό -- όπως ο προγραμματισμός ή η εφαρμογή της επιστημονικής μεθόδου -- ο μακροπρόθεσμος σχεδιασμός και ο αλγόριθμος

χειραγώγησης δεδομένων είναι εντελώς πέρα από αυτό που μπορούν να κάνουν οι τρέχουσες τεχνικές βαθιάς μάθησης, ακόμη και με μεγάλα δεδομένα.

Κεφάλαιο 8^ο

Επανάσταση Βαθιάς Μάθησης

Είναι βαθιά μάθηση ένα υποσύνολο της μηχανικής μάθησης και πώς η μηχανική μάθηση είναι ένα υποσύνολο της τεχνητής νοημοσύνης; (AI)

Το έτος 2012, μια ομάδα κέρδισε το "Merck Molecular Activity Challenge" έχοντας χρησιμοποιήσει νευρωνικά δίκτυα πολλαπλών εργασιών για να διαπιστώσει τον βιομοριακό στόχο ενός φαρμάκου. Το έτος 2014, μια άλλη η ομάδα χρησιμοποίησε τη βαθιά μάθηση για να εντοπίσει τόσο τις μη στοχευμένες όσο και τις τοξικές επιδράσεις των χημικών ουσιών σε θεραπευτικά συστατικά, προϊόντα οικιακής χρήσης και φάρμακα

Το 2011, αυτή η μέθοδος πέτυχε υπεράνθρωπα αποτελέσματα για πρώτη φορά στον διαγωνισμό οπτικής αναγνώρισης προτύπων. Το 2011 κέρδισε τον διαγωνισμό συγγραφής ICDAR της Κίνας. Τον Μάιο του 2012 κέρδισε τον Διαγωνισμό Τμηματοποίησης Εικόνας ISBI. Μόλις το 2011 έπαιξαν τα CNN Τον Ιούνιο του 2012, μια εργασία που παρουσιάστηκε στο κορυφαίο συνέδριο CVPR από τους Ciresan et al., δείχνει πώς η μέγιστη συγκέντρωση CNN στις GPU μπορεί να βελτιώσει σημαντικά πολλές εγγραφές οπτικής αναφοράς. Ένα παρόμοιο σύστημα από τους Krizhevsky et al., Οκτώβριος 2012. Οι μέθοδοι ρηχής μηχανικής μάθησης κέρδισαν τον μεγάλης κλίμακας διαγωνισμό ImageNet με σαφή διαφορά. Νοέμβριος 2012, The system of Ciresan et al. Κέρδισε επίσης τον διαγωνισμό ICPR για ανάλυση ιατρικής εικόνας μεγάλης κλίμακας για την ανίχνευση καρκίνου και το MICCAI Grand Challenge για το ίδιο θέμα την επόμενη χρονιά. Το 2013 και το 2014, παρόμοια με τις τάσεις αναγνώρισης ομιλίας μεγάλης κλίμακας, το ποσοστό σφάλματος για εργασίες ImageNet που χρησιμοποιούν βαθιά εκμάθηση μειώθηκε περαιτέρω.

Η ταξινόμηση εικόνων επεκτάθηκε στη συνέχεια στο πιο δύσκολο έργο της δημιουργίας περιγραφών εικόνων (λεζάντες). Ορισμένοι ερευνητές λένε ότι η νίκη του ImageNet τον Οκτώβριο του 2012 σηματοδοτεί την αρχή μιας «επανάστασης βαθιάς μάθησης» που θα μεταμορφώσει τη βιομηχανία τεχνητής νοημοσύνης. Τον Μάρτιο του 2019, στους Yoshua Bengio, Geoffrey Hinton και Yann LeCun απονεμήθηκε το βραβείο Turing για τις εννοιολογικές και μηχανολογικές τους ανακαλύψεις που έχουν κάνει τα βαθιά νευρωνικά δίκτυα σημαντικό δομικό στοιχείο της πληροφορικής.

Από το 2010, η πρόοδος στη μηχανική μάθηση και στο υλικό των υπολογιστών οδήγησε σε αποτελεσματικότερους τρόπους εκπαίδευσης νευρωνικών δικτύων. Μέχρι και το 2019, οι μονάδες επεξεργασίας γραφικών έχουν αντικαταστήσει τις CPU ως την κύρια μέθοδο εκπαίδευσης μεγάλης κλίμακας εμπορικής τεχνητής νοημοσύνης cloud.

Ειδικά ηλεκτρονικά κυκλώματα που ονομάζονται επεξεργαστές βαθιάς μάθησης σχεδιάστηκαν για να επιταχύνουν τους αλγόριθμους βαθιάς μάθησης. Οι επεξεργαστές βαθιάς εκμάθησης περιλαμβάνουν μονάδες νευρωνικής επεξεργασίας (NPU) σε κινητά τηλέφωνα Huawei και διακομιστές υπολογιστικού νέφους, όπως μονάδες επεξεργασίας τανυστών (TPU) στην πλατφόρμα Google Cloud. Η Cerebras Systems έχει επίσης κατασκευάσει ένα αποκλειστικό σύστημα για το χειρισμό μεγάλων μοντέλων βαθιάς εκμάθησης, το CS-2, που βασίζεται στον μεγαλύτερο επεξεργαστή στη βιομηχανία, τη δεύτερη γενιά Wafer Scale Engine (WSE-2).

Οι ατομικά λεπτοί ημιαγωγοί θεωρούνται πολλά υποσχόμενοι για ενεργειακά αποδοτικό υλικό βαθιάς εκμάθησης, όπου η ίδια βασική δομή συσκευής χρησιμοποιείται τόσο για λογικές λειτουργίες όσο και για αποθήκευση δεδομένων. Το 2020, οι Marega et al. Δημοσίευσε πειράματα με υλικό ενεργού καναλιού μεγάλης περιοχής για την ανάπτυξη συσκευών και κυκλωμάτων λογικής μνήμης που βασίζονται σε τρανζίστορ πεδίου αιωρούμενης πύλης (FGFET).

Το 2021, οι J. Feldmann et al. πρότεινε έναν ενσωματωμένο φωτονικό επιταχυντή υλικού για παράλληλη συνελκτική επεξεργασία. Οι συγγραφείς εντοπίζουν δύο βασικά πλεονεκτήματα της ολοκληρωμένης φωτονικής έναντι των ηλεκτρονικών ομολόγων της: (1) μαζικά παράλληλη μεταφορά δεδομένων μέσω πολυπλεξίας διαίρεσης μήκους κύματος σε συνδυασμό με χτένες συχνότητας και (2) εξαιρετικά υψηλές ταχύτητες διαμόρφωσης δεδομένων. Το σύστημά τους μπορεί να εκτελέσει τρισεκατομμύρια λειτουργίες πολλαπλασιασμού-συσσώρευσης ανά δευτερόλεπτο, υποδεικνύοντας τις δυνατότητες της ολοκληρωμένης φωτονικής σε εφαρμογές τεχνητής νοημοσύνης με μεγάλο όγκο δεδομένων.

Η βαθιά μάθηση έχει στενή σχέση με μια κατηγορία από θεωρίες για την ανάπτυξη του εγκεφάλου (πιο συγκεκριμένα την ανάπτυξη του νεοφλοιού), που παρουσιάστηκαν με τη μορφή υπολογιστικών μοντέλων που ήταν πρόδρομοι σε συστήματα βαθιάς μάθησης. Ένα κοινό νήμα μεταξύ αυτών των μοντέλων είναι οι διάφορες δυναμικές

μάθησης που προτείνονται στον εγκέφαλο, όπως η αύξηση των παραγόντων ανάπτυξης των νευρών, η υποστήριξη της αυτοοργάνωσης, κάπως παρόμοια με τα νευρωνικά δίκτυα που χρησιμοποιούνται στα μοντέλα βαθιάς μάθησης. Όπως ο νεοφλοιός, τα νευρωνικά δίκτυα χρησιμοποιούν μια ιεραρχία στρωμάτων φιλτραρίσματος, όπου κάθε στρώμα εξετάζει πληροφορίες από το προηγούμενο στρώμα (ή λειτουργικό περιβάλλον) και στη συνέχεια περνά την έξοδο του (και πιθανώς την αρχική είσοδο) σε άλλα επίπεδα. Αυτή η διαδικασία παράγει μια αυτό-οργανωμένη στοίβα αισθητήρων που είναι καλά προσαρμοσμένη στο περιβάλλον στο οποίο λειτουργούν. Μια περιγραφή του 1995 αναφέρει: «...ο εγκέφαλος του βρέφους φαίνεται να είναι οργανωμένος υπό την επίδραση κυμάτων των λεγόμενων τροφικών παραγόντων... διαφορετικές περιοχές του εγκεφάλου συνδέονται διαδοχικά, με το ένα στρώμα ιστού να ωριμάζει πριν από το άλλο, έτσι ώστε ο συνολικός ώριμος εγκέφαλος».

Διάφορες προσεγγίσεις έχουν χρησιμοποιηθεί για τη διερεύνηση της εγκυρότητας των μοντέλων βαθιάς μάθησης από νευροβιολογική προοπτική. Από τη μία πλευρά, έχουν προταθεί αρκετές παραλλαγές του αλγορίθμου backpropagation για να αυξηθεί ο ρεαλισμός της επεξεργασίας του. Άλλοι ερευνητές πιστεύουν ότι οι μη εποπτευόμενες μορφές βαθιάς μάθησης, όπως αυτές που βασίζονται σε ιεραρχικά παραγωγικά μοντέλα και δίκτυα βαθιάς πεποίθησης, μπορεί να είναι πιο κοντά στη βιολογική πραγματικότητα. Από αυτή την άποψη, τα μοντέλα γενετικών νευρωνικών δικτύων συνδέονται με νευροβιολογικά στοιχεία για επεξεργασία βάσει δειγματοληψίας στον εγκεφαλικό φλοιό.

Δεν έχει ακόμη καθιερωθεί συστηματική σύγκριση μεταξύ της νευρωνικής κωδικοποίησης στον ανθρώπινο εγκεφαλικό ιστό και των βαθιών δικτύων, αλλά έχουν αναφερθεί ορισμένες αναλογίες. Για παράδειγμα, οι υπολογισμοί που εκτελούνται από μονάδες βαθιάς μάθησης μπορεί να μοιάζουν με αυτούς των πραγματικών νευρώνων και των νευρωνικών πληθυσμών. Ομοίως, οι αναπαραστάσεις που αναπτύσσονται από μοντέλα βαθιάς μάθησης είναι παρόμοιες με εκείνες που μετρούνται από το οπτικό σύστημα των πρωτευόντων σε επιμέρους επίπεδα μονάδας και πληθυσμού.

Κριτική και σχόλιο

Η βαθιά μάθηση έχει να αντιμετωπίσει κριτικές και σχόλια, σε κάποιες περιπτώσεις ακόμη και πέρα από τον τομέα της επιστήμης των υπολογιστών. Αυτό ονομάζεται «αντίπαλη επίθεση».

Μια σημαντική κριτική είναι η έλλειψη θεωρίας γύρω από κάποιες μεθόδους. Η μάθηση στις πιο κοινές αρχιτεκτονικές χρησιμοποιεί την καλά κατανοητή βαθμίδωση. Η θεωρία που περιλαμβάνει άλλους αλγόριθμους όπως η αντίθεση απόκλισης δεν είναι σαφής. Οι μέθοδοι της βαθιάς μάθησης αντιμετωπίζονται κυρίως ως μαύρα κουτιά και οι περισσότερες επιβεβαιώσεις είναι εμπειρικές και όχι θεωρητικές. Επισημάνθηκε επίσης ότι η βαθιά μάθηση πρέπει να θεωρείται ως ένα σημαντικό βήμα προς την ισχυρή τεχνητή νοημοσύνη και όχι ως μια ολοκληρωμένη λύση. Όσο ισχυρές κι αν είναι οι μέθοδοι βαθιάς μάθησης, εξακολουθούν να μην έχουν πολλά από τα χαρακτηριστικά που είναι απαραίτητα για την πλήρη ολοκλήρωση αυτού του στόχου. Ένας ερευνητής ψυχολόγος σημειώνει:

«Στην πραγματικότητα, η βαθιά μάθηση είναι μ ένα μέρος μιας μεγαλύτερης πρόκλησης της κατασκευής μηχανών . Τέτοιες τεχνολογίες δεν αναπαριστούν αιτιατές σχέσεις, ούτε έχουν κάλους τρόπους εκτέλεσης λογικού συλλογισμού και επίσης απέχουν πολύ από την διαμόρφωση αφηρημένης γνώσης, δηλαδή δεδομένα σχετικά με αντικείμενο είναι, σε τι χρησιμεύουν και πώς αυτά χρησιμοποιούνται.»

Αναφερόμενοι περαιτέρω στην ιδέα ότι οι καλλιτεχνικές ευαισθησίες μπορεί να είναι εγγενείς σε χαμηλά επίπεδα γνώσης, δημοσιευμένες αναπαραστάσεις της εσωτερικής κατάστασης των νευρωνικών δικτύων βαθιάς (20-30 επίπεδα) επιχειρούν να διακρίνουν τις εικόνες μεταξύ ουσιαστικά τυχαίων δεδομένων στα οποία παρουσιάζονται καλά και οπτικά ελκυστικό: το πρωτότυπο σημείωμα έρευνας έλαβε πάρα πολλά σχόλια και ήταν ο θέμα του άρθρου με τις περισσότερες επισκέψεις στον ιστότοπο Guardian για αρκετό καιρό.

Σφάλματα

Κάποιες από τις αρχιτεκτονικές βαθιάς μάθησης παρουσιάζουν προβληματική συμπεριφορά, όπως η ταξινόμηση των μη αναγνωρισμένων εικόνων που ανήκουν σε γνωστές κοινές κατηγορίες εικόνων και η λάθος ταξινόμηση ν διαταραχών των σωστά

ταξινομημένων εικόνων. Υποθέτουμε ότι οι συμπεριφορές αυτές έχουν σχέση με τους περιορισμούς της εσωτερικής του αναπαράστασης και ότι οι περιορισμοί αυτοί είναι που εμποδίζουν την ενσωμάτωση σε διαφορετικές αρχιτεκτονικές τεχνητής νοημοσύνης πολλαπλών συστατικών. Τα συγκεκριμένα προβλήματα μπορούν να διορθωθούν από τις εσωτερικές συγγενείς καταστάσεις που σχηματίζουν αρχιτεκτονικές βαθιάς μάθησης για την επίλυση της σύνταξης του γραφήματος των παρατηρούμενων οντοτήτων.

Κυβερνοαπειλή

Όσο η βαθιά μάθηση μεταδίδεται από το εργαστήριο στον κόσμο, η έρευνα καθώς και η εμπειρία έχουν δείξει ότι τα νευρωνικά δίκτυα είναι εύκολο να πλαστογραφηθούν. Αναγνωρίζοντας τα χαρακτηριστικά που χρησιμοποιούν αυτά τα συστήματα ώστε να λειτουργήσουν σωστά, ένας εισβολέας μπορεί και τροποποιήσει την είσοδο στο τεχνητό νευρωνικό δίκτυο, αναγκάζοντας το τεχνητό νευρωνικό δίκτυο να βρει αντιστοιχίες που ένας ανθρώπινος παρατηρητής δεν θα αναγνώριζε. Ένας εισβολέας θα μπορούσε να προβεί σε ανεπαίσθητες τροποποιήσεις σε μια εικόνα, έτσι ώστε ένα τεχνητό νευρωνικό δίκτυο να μην μπορεί να βρει μια ομοιότητα, ακόμα κι αν η εικόνα δεν μοιάζει με αυτό που θα αναζητούσε ένας άνθρωπος. αυτή η χειραγώγηση.

Το 2016, ερευνητές χρησιμοποίησαν τεχνητό νευρωνικό δίκτυο για να επεξεργαστούν κάποιες εικόνες, εντοπίζοντας σε τι εστιάζουν οι άλλοι άνθρωποι και δημιουργώντας εικόνες που τους ξεγέλασαν. Η τροποποιημένη εικόνα δεν φαίνεται καθόλου διαφορετική στο ανθρώπινο μάτι. Μια άλλη ομάδα έδειξε ότι η επακόλουθη εκτύπωση μορφοποιημένων εικόνων που ελήφθησαν ξεγέλασε με επιτυχία τα συστήματα που ταξινομούν εικόνες. Μια ερευνά έδειξε ότι ορισμένα εργαλεία μπορούν να μπερδέψουν τα συστήματα που αναγνωρίζουν πρόσωπα και αντιμετωπίζουν τους ανθρώπους ως διασημότητες, και επιτρέπουν δυνητικά σε ένα άτομο να υποδυθεί ένα άλλο.

Ωστόσο, τα τεχνητά νευρωνικά δίκτυα μπορούν να εκπαιδευτούν περαιτέρω για να ανιχνεύουν απόπειρες πλαστογράφησης, οι οποίες θα μπορούσαν να οδηγήσουν επιτιθέμενους και υπερασπιστές σε μια κούρσα εξοπλισμών παρόμοια με αυτή που έχει ορίσει τη βιομηχανία αμυντικού κακόβουλου λογισμικού. Το τεχνητό νευρωνικό δίκτυο εκπαιδεύεται για να νικήσει το anti-malware που βασίζεται σε τεχνητό

νευρωνικό δίκτυο επιτιθέμενος επανειλημμένα σε άμυνες κακόβουλου λογισμικού που τροποποιούνται συνεχώς από γενετικούς αλγόριθμους

Το 2016, μια άλλη ομάδα έδειξε ότι ορισμένοι ήχοι θα μπορούσαν να κάνουν το σύστημα φωνητικών εντολών Google Now να ανοίξει μια συγκεκριμένη διεύθυνση ιστού και υπέθεσε ότι αυτό θα μπορούσε να «χρησιμοποιήσει ως σκαλοπάτι για περαιτέρω επιθέσεις (π.χ. άνοιγμα μιας ιστοσελίδας που φιλοξενεί κακόβουλο λογισμικό Drive-by). Στη «δηλητηρίαση δεδομένων», λανθασμένα δεδομένα εισάγονται συνεχώς στο εκπαιδευτικό σετ ενός συστήματος μηχανικής μάθησης για να αποτραπεί η κατάκτησή του.

Εξάρτηση από τον άνθρωπο

Τα συστήματα βαθιάς μάθησης βασίζονται σε δεδομένα της εκπαίδευσης και επικύρωσης που έχουν δημιουργηθεί από τον άνθρωπο και έχουν σχολιαστεί. Στη φιλοσοφία των μέσων ενημέρωσης, έχει υποστηριχθεί ότι όχι μόνο αναπτύσσονται συχνά για το σκοπό αυτό εργασίες χαμηλών αμειβόμενων κλικ (όπως στο Amazon Mechanical Turk), αλλά έχουν αναπτυχθεί και άλλες μορφές ανθρώπινης εργασίας που συχνά δεν αναγνωρίζονται. Υπάρχουν πέντε τύποι μηχανικής της ανθρώπινης εργασίας για τη δημιουργία δεδομένων εκπαίδευσης:

1. παιχνιδιοποίηση
2. "σύλληψη και παρακολούθηση"
3. αξιοποίηση κοινωνικών κινήτρων
4. εξόρυξη πληροφοριών και
5. clickwork.

Στις περισσότερες εμπορικές εφαρμογές τελικού χρήστη βαθιάς μάθησης, η ζήτηση για δεδομένα εκπαίδευσης δεν σταματά. Αντίθετα, υπάρχει συνεχής ζήτηση για δεδομένα επικύρωσης τα οποία παράγονται από τον άνθρωπο για τη συνεχή και ενημέρωση των τεχνητών νευρωνικών δικτύων. Για το λόγο αυτό, η πλατφόρμα του Facebook χρησιμοποιεί μια λειτουργία όπου οι χρήστες θα ειδοποιούνται όταν τακτοποιούνται σε μια εικόνα. Έτσι μπορούν οι ίδιοι να αποφασίσουν αν δεν τους αρέσει να επισημαίνονται ή ενημερώσουν την πλατφόρμα ότι δεν είναι αυτοί.

Η εμπλοκή των χρηστών στη δημιουργία δεδομένων εκπαίδευσης και επικύρωσης είναι χαρακτηριστική των περισσότερων εμπορικών εφαρμογών τελικού χρήστη βαθιάς μάθησης, έτσι ώστε τέτοια συστήματα μπορούν να ονομαστούν "τεχνητή νοημοσύνη υποβοηθούμενη από τον άνθρωπο".

ΚΕΦΑΛΑΙΟ 9^ο

Μηχανική Μάθηση

Σε αυτό και το επόμενο κεφάλαιο, παρουσιάζεται η θεωρία του υπόβαθρου της Μηχανικής Μάθησης που χρησιμοποιήθηκε στην παρούσα εργασία, καθώς και των Συνελκτικών Νευρωνικών Δικτύων, αρχιτεκτονικές και εφαρμογές τους με την πρόθεση να αναλυθεί το σύστημα αναγνώρισης φαγητού σε εικόνες.

Ορισμός

Μηχανικά μάθηση ονομάζουμε το κομμάτι της επιστήμης υπολογιστών που προσπαθεί να μοντελοποιήσει πολλαπλά επίπεδα αφαιρετικότητας, χρησιμοποιώντας πολλαπλά επίπεδα στις αρχιτεκτονικές της. Τα δίκτυα αυτά είναι εμπνευσμένα από το πώς επεξεργάζεται ο άνθρωπος την πληροφορία και προσπαθούν να προσομοιάσουν την λειτουργία των νευρώνων στο νεοφλοιό του εγκεφάλου (όπου γίνεται περίπου το 80% της ανθρώπινης σκέψης). Η διαδικασία της μάθησης ξεκινά με παρατηρήσεις, που αποτελούν παραδείγματα, ή εμπειρικά αποτελέσματα ή οδηγίες, ούτως ώστε να αναγνωριστούν πρότυπα στα δεδομένα και να ληφθούν καλύτερες αποφάσεις στο μέλλον, με βάση τα παραδείγματα που διαθέτουμε. Ο πρωταρχικός σκοπός είναι να επιτρέψουμε στους υπολογιστές να μαθαίνουν αυτόματα, χωρίς ανθρώπινη παρέμβαση ή βοήθεια, και να προσαρμόζουν τις πράξεις τους κατάλληλα.

Στη μηχανική μάθηση, τα καθήκοντα ταξινομούνται γενικά σε ευρείες κατηγορίες. Οι κατηγορίες αυτές βασίζονται στον τρόπο με τον οποίο λαμβάνεται η μάθηση ή στον τρόπο με τον οποίο δίνεται ανάδραση στην εκμάθηση στο ανεπτυγμένο σύστημα. Δύο από τις πιο ευρέως υιοθετημένες μεθόδους είναι η επιβλεπόμενη μάθηση, η οποία εκπαιδεύει αλγόριθμους που βασίζονται στα δεδομένα εισόδου και εξόδου τα οποία επισημαίνονται (αποκτούν ετικέτες-labels) από τον άνθρωπο και η μη επιβλεπόμενη μάθηση, η οποία παρέχει τον αλγόριθμο χωρίς επισημασμένα δεδομένα, ούτως ώστε να του επιτρέψει να βρει δομή στα δεδομένα εισόδου του.

Μάθηση με Επίβλεψη

Η Μάθηση με Επίβλεψη είναι ο τομέας μηχανικής μάθησης όπου μια συνάρτηση μαθαίνει να αντιστοιχίζει δεδομένα εισόδου σε δεδομένα εξόδου χρησιμοποιώντας παραδείγματα ζευγών εισόδου-εξόδου. Αυτό το σύνολο ζευγών ονομάζεται σύνολο

εκπαίδευσης και η διαδικασία υπολογισμού μιας τέτοιας συνάρτησης από το παραπάνω σύνολο λέγεται εκπαίδευση. Ο σκοπός της μάθησης με επίβλεψη είναι ο υπολογισμός μιας συνάρτησης που γενικεύει επαρκώς σε δεδομένα εισόδου στα οποία δεν έχει εκπαιδευτεί, αντιστοιχίζοντάς τα σε σωστές εξόδους.

Η μάθηση με επίβλεψη είναι κατάλληλη για εργασίες όπου, ενώ η απεικόνιση εισόδου εξόδου είναι δύσκολο-πολύπλοκο να βρεθεί αναλυτικά, υπάρχει ένα αρκετά μεγάλο σύνολο δεδομένων εκπαίδευσης. Μια τέτοια περίπτωση είναι η ταξινόμηση εικόνων. Από προγραμματιστική άποψη, οι εικόνες αναπαριστώνται από 3D τένσορες με διαστάσεις ύψους, πλάτους και βάθους (καναλιών, πχ RGB). Ένα σύνολο εκπαίδευσης για ταξινόμηση εικόνων περιέχει μια συλλογή εικόνων και αντίστοιχες επισημάνσεις. Ένα παράδειγμα είναι εικόνες με αυτοκίνητα και ποδήλατα με τις αντίστοιχες επισημάνσεις "αυτοκίνητο" ή "ποδήλατο" για κάθε μια από αυτές.

Στη μάθηση με επίβλεψη, οι είσοδοι του συνόλου εκπαίδευσης πρέπει να μετατραπούν σε καταλληλότερες δομές ώστε να γίνει περαιτέρω επεξεργασία. Τέτοια είναι η περίπτωση για σύνολα εκπαίδευσης σχετικά με κείμενο και η κλασική προσέγγιση στην όραση υπολογιστών. Πριν τα συνελκτικά νευρωνικά δίκτυα γίνουν δημοφιλή, η διαδικασία ταξινόμησης εικόνων περιλάμβανε την μετατροπή κάθε εικόνας σε ένα σύνολο διανυσμάτων περιγραφής της εικόνας, τα χαρακτηριστικά, τα οποία χρησιμοποιούνταν σαν είσοδος για την εκπαίδευση μοντέλων. Ένα τυπικό παράδειγμα χαρακτηριστικών είναι οι γωνίες. Η διαδικασία σχεδιασμού τέτοιων χαρακτηριστικών περιλάμβανε κοπιαστική παραμετροποίηση, ενώ τα αποτελέσματα αυτού του πλαισίου εργασίας φάνηκε να μην βελτιώνεται περαιτέρω προς το τέλος της δεκαετίας του 2000. Με την εισαγωγή των ΣΝΔ, η διαδικασία σχεδιασμού χαρακτηριστικών έγινε μέρος της διαδικασίας εκπαίδευσης. Τα ΣΝΔ υπολογίζουν τα δικά τους, ειδικά χαρακτηριστικά, τα οποία δεν είναι διαισθητικά κατανοητά από ανθρώπους, όπως είναι οι γωνίες.

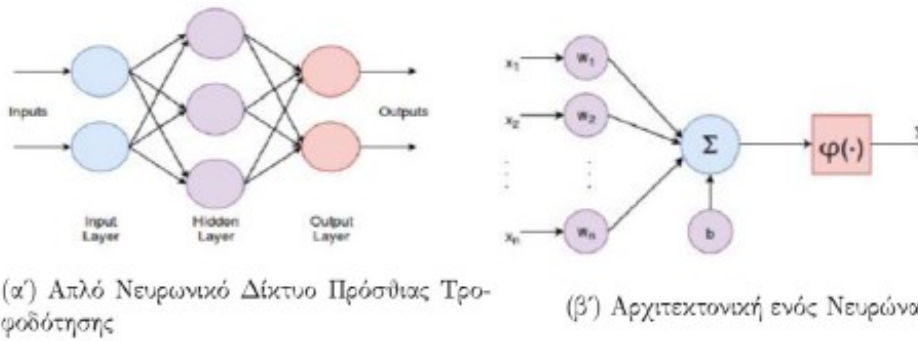
Μια ακόμα εργασία Μάθησης με Επίβλεψη, εκτός της ταξινόμησης, είναι η παλινδρόμηση, όπου, αντί για κατηγορία ως έξοδο, αναμένεται αριθμητική τιμή. Στο πλαίσιο της ανίχνευσης αντικειμένων από εικόνες, ένα παράδειγμα είναι να βρεθούν οι συντεταγμένες που ορίζουν ένα ορθογώνιο περίγραμμα γύρω από το αυτοκίνητο.

Μάθηση χωρίς Επίβλεψη

Σε άλλα προβλήματα μηχανικής μάθησης, υπάρχει μια διαφορετική τάξη καθηκόντων που αναφέρεται ως μη επιβλεπόμενη μάθηση. Στα προβλήματα αυτής της κατηγορίας, τα δεδομένα εκπαίδευσης είναι διανύσματα x τα οποία δεν έχουν αντίστοιχες ετικέτες. Επομένως, ο στόχος τη μη επιβλεπόμενης μάθησης είναι να βρίσκει μοτίβα όταν δεν υπάρχουν “σωστές απαντήσεις”, ή όταν αυτές είναι αδύνατον να υπολογιστούν. Μία μεγάλη υποκατηγορία μη επιβλεπόμενων τεχνικών είναι το πρόβλημα της ομαδοποίησης (clustering). Η μέθοδος αυτή αναφέρεται στην ομαδοποίηση παρατηρήσεων με τέτοιο τρόπο ούτως ώστε τα μέλη μιας κοινής ομάδας να είναι παρόμοια το ένα με το άλλο, και να διαφέρουν σημαντικά από τα μέλη των άλλων ομάδων. Μια άλλη πολύ ενδιαφέρουσα κατηγορία μη επιβλεπόμενων καθηκόντων είναι οι τα γεννητικά μοντέλα (generative models). Τα μοντέλα αυτά μιμούνται τη διαδικασία δημιουργίας των δεδομένων εκπαίδευσης. Ένα καλό γεννητικό μοντέλο θα πρέπει να μπορεί να δημιουργήσει νέα δεδομένα τα οποία, αν και είναι τεχνητά, μοιάζουν με τα αυθεντικά. Αυτός ο τρόπος μάθησης είναι μη επιβλεπόμενος διότι η διαδικασία με την οποία δημιουργούνται (“γεννιούνται”) τα δεδομένα δεν είναι άμεσα παρατηρήσιμη - μόνο τα ίδια τα δεδομένα είναι παρατηρήσιμα.

Νευρωνικά Δίκτυα

Ένα παράδειγμα συνάρτησης Μάθησης με Επίβλεψη είναι τα Νευρωνικά Δίκτυα Πρόσθιας Τροφοδότησης (ΝΔΠΤ). Τα νευρωνικά αυτά δίκτυα αποτελούνται από ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου, σειριακά συνδεδεμένα. Το επίπεδο εισόδου παρέχει ένα διάνυσμα εισόδου στο δίκτυο, ενώ το επίπεδο εξόδου παρέχει την πρόβλεψη σαν μια τιμή ή διάνυσμα. Τα κρυφά επίπεδα ορίζουν την πολύπλοκη εσωτερική λειτουργία του δικτύου. Η γενική όψη τους παρουσιάζεται στο παρακάτω σχήμα:



Η δομή των νευρωνικών δικτύων πρόσθιας τροφοδότησης είναι ένας κατευθυνόμενος ακυκλικός γράφος, όπου κάθε κόμβος, ο Νευρώνας, είναι μια μη γραμμική συνάρτηση $R^n \rightarrow R$. Ο Νευρώνας υπολογίζει την έξοδο σε δύο βήματα - υπολογίζει το σταθμισμένο άθροισμα των n εισόδων του συν έναν όρο πόλωσης και στη συνέχεια εφαρμόζει μια συνάρτηση κανονικοποίησης $\varphi(\cdot)$ στο άθροισμα. Η τελευταία λέγεται συνάρτηση ενεργοποίησης. Υπάρχουν $n + 1$ παράμετροι σχετικές του κάθε Νευρώνα, τα n βάρη w_1, \dots, w_n και η πόλωση b , όπως φαίνεται κι από την παρακάτω εξίσωση. Η έξοδος κάθε νευρώνα γίνεται είσοδος σε κάθε Νευρώνα στο επόμενο επίπεδο. Το νευρωνικό δίκτυο πρόσθιας τροφοδότησης μπορεί να ρυθμιστεί με εκπαίδευση με προσαρμογή του συνόλου των παραμέτρων των νευρώνων.

$$y = \varphi \left(\sum_{j=0}^n w_j x_j + b \right)$$

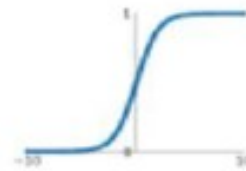
Χρησιμοποιώντας μια μη γραμμική συνάρτηση ενεργοποίησης, το νευρωνικό δίκτυο πρόσθιας τροφοδότησης που προκύπτει γίνεται επίσης μη γραμμικό. Παραδείγματα συναρτήσεων ενεργοποίησης είναι η Σιγμοειδής (Sigmoid), η Υπερβολική Εφαπτομένη (Tanh) και η ReLU. Οι μη γραμμικές συναρτήσεις ενεργοποίησης κάνουν τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης καθολικές προσεγγίσεις, δηλαδή μπορούν να προσεγγίσουν επαρκώς συνεχείς συναρτήσεις ορισμένες σε συμπαγή υποσύνολα του R^n επιλέγοντας κατάλληλες παραμέτρους και έναν επαρκή, αλλά πεπερασμένο, αριθμό νευρώνων. Αυτή η ιδιότητα προσθέτει μια θεωρητική δικαιολόγηση χρήσης νευρωνικού δικτύου πρόσθιας τροφοδότησης για εργασίες Μάθησης με Επίβλεψη.

Η εκπαίδευση νευρωνικού δικτύου πρόσθιας τροφοδότησης γίνεται με τον δημοφιλή αλγόριθμο οπισθοδιάδοσης. Η οπισθοδιάδοση είναι ένας αλγόριθμος βελτιστοποίησης που ψάχνει για παραμέτρους (βάρη και πολώσεις) που ελαχιστοποιούν την συνάρτηση σφάλματος. Το σφάλμα ορίζεται ως μια συνάρτηση που μετράει την απόσταση (σφάλμα) μεταξύ των εξόδων του συνόλου εκπαίδευσης και των αντίστοιχων εξόδων υπολογισμένων από το νευρωνικό δίκτυο πρόσθιας τροφοδότησης χρησιμοποιώντας τις εισόδους του συνόλου εκπαίδευσης. Η διαισθητική επεξήγηση της οπισθοδιάδοσης είναι ότι το σφάλμα υπολογίζεται στην έξοδο και διανέμεται προς τα προηγούμενα επίπεδα του δικτύου, ανανεώνοντας τις παραμέτρους ανάλογα με την ευαισθησία τους σε αλλαγές του σφάλματος, χρησιμοποιώντας τον κανόνα παραγωγικής αλυσίδας.

Παρακάτω φαίνονται μερικές συναρτήσεις ενεργοποίησης:

Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



tanh

$\tanh(x)$



ReLU

$\max(0, x)$



Optimizers

Έστω ότι η συνάρτηση που πρέπει να ελαχιστοποιηθεί είναι η $f(x)$. Τότε η κλίση της θα είναι η $\Delta f(x)$. Το μέγεθος του βήματος για την k επανάληψη είναι t_k . Παρακάτω αναλύονται οι πιο διαδεδομένοι αλγόριθμοι ελαχιστοποίησης της συνάρτησης.

Batch Gradient Descent

Ο αλγόριθμος Batch Gradient Descent ενημερώνει τις παραμέτρους x διατρέχοντας όλο το σύνολο δεδομένων ως εξής:

$$x_{k+1} = x_k - t_k \Delta f(x_k)^{(1:n)}$$

Ο αλγόριθμος αυτός εγγυάται σύγκλιση σε ολικό ελάχιστο για κυρτό πρόβλημα (convex problem) και σύγκλιση σε τοπικό ελάχιστο για μη-κυρτό πρόβλημα (non-convex problem). Όμως σε σύγχρονα προβλήματα βαθιάς μάθησης θα πάρει πολύ χρόνο και μνήμη ο υπολογισμός του ολικού ελαχίστου από ολόκληρο το σύνολο δεδομένων. Οπότε η μέθοδος αυτή ελαχιστοποίησης χρησιμοποιείται σπάνια σε προβλήματα βαθιάς μάθησης.

Stochastic Gradient Decent

Αντίθετα, ο αλγόριθμος Stochastic Gradient Decent υπολογίζει τις κλίσεις και ενημερώνει τις παραμέτρους για κάθε δεδομένο του συνόλου εκπαίδευσης:

$$x_{k+1} = x_k - t_k \Delta f(x_k)^{(i)}$$

Όμως, λόγω της διακύμανσης (variance) των δεδομένων του συνόλου εκπαίδευσης, η αντικειμενική συνάρτηση (objective function) θα διακυμαίνεται έντονα. Παρόλο που το μικρό μέγεθος βήματος επιτρέπει στον αλγόριθμο να συγκλίνει σε ένα καλό σημείο, η εκπαίδευση είναι αργή.

Mini-Batch Gradient Decent

Ο Mini-Batch Gradient Decent συνδυάζει τα πλεονεκτήματα των προηγούμενων δύο αλγορίθμων και ενημερώνει τις παραμέτρους αφού έχει υπολογίσει την κλίση σε ένα μικρό σύνολο (batch) δεδομένων:

$$x_{k+1} = x_k - t_k \Delta f(x_k)^{(i:i+m)}$$

όπου m είναι το μέγεθος του μικρού αυτού συνόλου. ο Mini-Batch Gradient Decent δεν εγγυάται σύγκλιση σε καλό σημείο και το μέγεθος βήματος απαιτεί εμπειρία. Οπότε οι ερευνητές επεκτείνουν τον αλγόριθμο αυτό για να πετύχουν καλύτερο σημείο σύγκλισης.

Adagrad

Ο Adagrad προσαρμόζει το μέγεθος του βήματος για κάθε παράμετρο ανάλογα με το ιστορικό ενημερώσεων της παραμέτρου:

$$G_k = G_{k-1} + \Delta f(x_k)^2$$
$$x_{k+1} = x_k - \frac{t}{\sqrt{G_k + \varepsilon}} \Delta f(x_k)$$

όπου G είναι μια συσσώρευση (accumulation) του ιστορικού της παραμέτρου και ε παράμετρος απάλυνσης (smoothing) για αποφυγή διαίρεσης με το μηδέν και είναι της τάξης του $1e-6$. Το μέγεθος βήματος είναι μεγαλύτερο για τις παραμέτρους με μικρό G και μεγάλο για παραμέτρους με μεγάλο G . Όμως, όταν το G μεγαλώσει αρκετά, το μέγεθος του βήματος θα τείνει στο μηδέν ύστερα από μεγάλο πλήθος επαναλήψεων. Οπότε έχουν προταθεί οι επόμενες μέθοδοι.

Adadelta

Ο Adadelta προέρχεται από τον Adagrad και βελτιώνει τα κύρια δύο μειονεκτήματά του:

1. Τη συνεχή φθορά(decay) των βαθμών εκμάθησης(learning rates)
2. Την ανάγκη για χειρονακτική επιλογή καθολικού learning rate

Ο Adadelata κλιμακώνει το μέγεθος του βήματος σύμφωνα με την τελευταία ενημέρωση του ιστορικού σε αντίθεση με τον Adagrad που χρησιμοποιεί όλο το ιστορικό. Επίσης χρησιμοποιεί έναν όρο επιτάχυνσης (acceleration term):

$$\begin{aligned}\mathbb{E}[\Delta f(x)^2]_k &= \rho \mathbb{E}[\Delta f(x)^2]_{k-1} + (1 - \rho) \Delta f(x_k)^2 \\ \hat{x}_k &= - \frac{\sqrt{\mathbb{E}[\hat{x}^2]_{k-1} + \varepsilon}}{\sqrt{\mathbb{E}[\Delta f(x)^2]_k + \varepsilon}} \Delta f(x_k) \\ \mathbb{E}[\hat{x}^2]_k &= \rho \mathbb{E}[\hat{x}^2]_{k-1} + (1 - \rho) \hat{x}_k^2 \\ x_{k+1} &= x_k + \hat{x}_k\end{aligned}$$

όπου ρ είναι μια σταθερά φθοράς(decay constant) και ε παράμετρος σταθερότητας(stability constant).

RMSprop

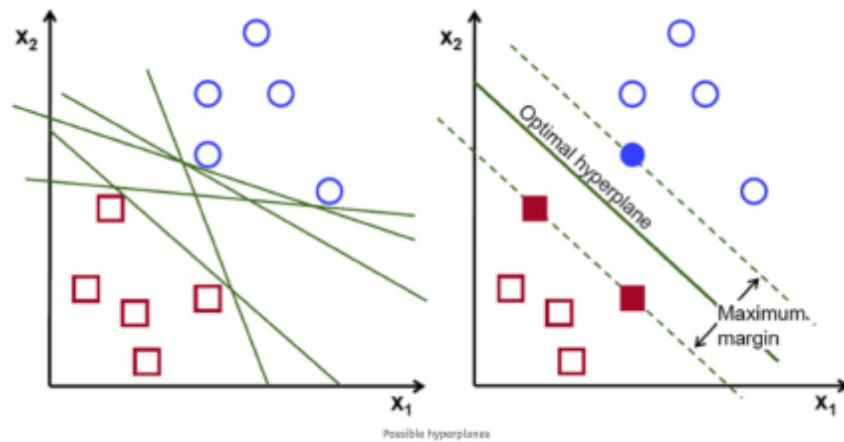
Ο RMSprop χρησιμοποιείται επίσης για την αντιμετώπιση των προβλημάτων του Adagrad:

$$\begin{aligned}\mathbb{E}[\Delta f(x)^2]_k &= \rho \mathbb{E}[\Delta f(x)^2]_{k-1} + (1 - \rho) \Delta f(x_k)^2 \\ x_{k+1} &= x_k - \frac{t}{\sqrt{\mathbb{E}[\Delta f(x)^2]_k + \varepsilon}} \Delta f(x_k)\end{aligned}$$

Γραμμική Ταξινόμηση

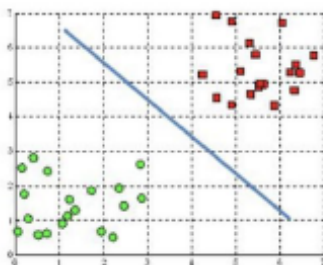
Μηχανές Διανυσμάτων Υποστήριξης

Ο στόχος του αλγόριθμου μηχανής διανυσματικής υποστήριξης[2] είναι να βρεθεί ένα υπερπλάνο σε ένα χώρο N -διαστάσεων (N - ο αριθμός των χαρακτηριστικών) που ταξινομεί τα σημεία δεδομένων.

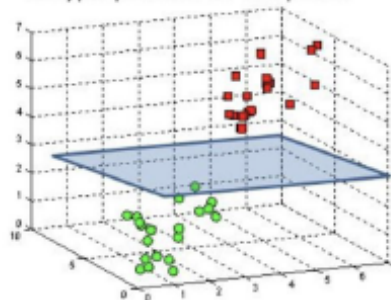


Για να διαχωρίσουμε τις δύο κατηγορίες σημείων δεδομένων, υπάρχουν πολλά πιθανά υπερπλάνα που θα μπορούσαν να επιλεγούν. Στόχος μας είναι να βρούμε ένα υπερπλάνο που έχει το μέγιστο περιθώριο, δηλαδή τη μέγιστη απόσταση μεταξύ σημείων δεδομένων και των δύο κατηγοριών. Η μεγιστοποίηση της απόστασης περιθωρίου παρέχει κάποια ενίσχυση, ώστε τα μελλοντικά σημεία δεδομένων να μπορούν να ταξινομηθούν με μεγαλύτερη αξιοπιστία.

A hyperplane in \mathbb{R}^2 is a line

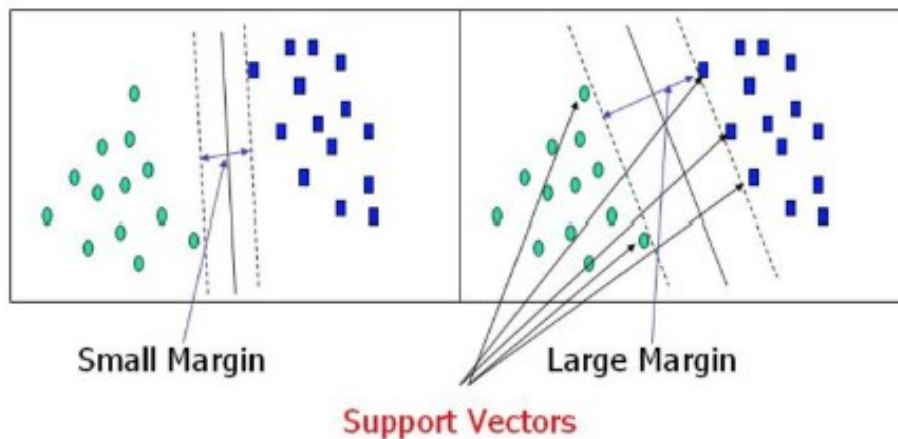


A hyperplane in \mathbb{R}^3 is a plane



Hyperplanes in 2D and 3D feature space

Τα υπερπλάνα είναι όρια απόφασης που βοηθούν στην ταξινόμηση των σημείων δεδομένων. Τα σημεία δεδομένων που εμπίπτουν σε κάθε πλευρά του υπερπλάνου μπορούν να αποδοθούν σε διαφορετικές κατηγορίες. Επίσης, η διάσταση του υπερπλάνου εξαρτάται από τον αριθμό των χαρακτηριστικών. Εάν ο αριθμός των χαρακτηριστικών εισόδου είναι 2, τότε το υπερπλάνο είναι μόνο μια γραμμή. Εάν ο αριθμός των χαρακτηριστικών εισόδου είναι 3, τότε το υπερπληρωμή γίνεται ένα δισδιάστατο επίπεδο. Είναι δύσκολο να φανταστεί κανείς πώς θα είναι το υπερπλάνο όταν ο αριθμός των χαρακτηριστικών ξεπερνά τα 3.



Τα διανύσματα υποστήριξης είναι σημεία δεδομένων που είναι πιο κοντά στο υπερπλάνο και επηρεάζουν τη θέση και τον προσανατολισμό του υπερπλάνου. Χρησιμοποιώντας αυτούς τους φορείς υποστήριξης, μεγιστοποιούμε το περιθώριο του ταξινομητή. Η διαγραφή των διανυσμάτων υποστήριξης θα αλλάξει τη θέση του υπερπλάνου. Αυτά είναι τα σημεία που μας βοηθούν να χτίσουμε το SVM μας.

Συνάρτηση κόστους και ενημέρωση κλίσεων

Στον αλγόριθμο SVM, επιδιώκουμε να μεγιστοποιήσουμε το περιθώριο μεταξύ των σημείων δεδομένων και του υπερπλάνου. Η συνάρτηση κόστους που βοηθά στη μεγιστοποίηση του περιθωρίου είναι:

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases} \quad c(x, y, f(x)) = (1 - y * f(x))_+$$

Hinge loss function (function on left can be represented as a function on the right)

Το κόστος είναι 0 εάν η προβλεπόμενη τιμή και η πραγματική τιμή είναι του ίδιου σημείου. Εάν δεν είναι, τότε υπολογίζουμε την τιμή της ζημίας. Προσθέτουμε επίσης μια παράμετρο κανονικοποίησης στη συνάρτηση κόστους. Ο στόχος της παραμέτρου κανονικοποίησης είναι η εξισορρόπηση της μεγιστοποίησης περιθωρίου και της απώλειας. Μετά την προσθήκη της παραμέτρου κανονικοποίησης, οι λειτουργίες κόστους φαίνονται όπως παρακάτω.

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^n (1 - y_i \langle x_i, w \rangle)_+$$

Loss function for SVM

Τώρα που έχουμε τη συνάρτηση κόστους, παίρνουμε μερικές παραγώγους σε σχέση με τα βάρη για να βρούμε τις κλίσεις. Χρησιμοποιώντας τις κλίσεις, μπορούμε να ενημερώσουμε τα βάρη μας.

$$\frac{\delta}{\delta w_k} \lambda \|w\|^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k} (1 - y_i \langle x_i, w \rangle)_+ = \begin{cases} 0, & \text{if } y_i \langle x_i, w \rangle \geq 1 \\ -y_i x_{ik}, & \text{else} \end{cases}$$

Gradients

Όταν δεν υπάρχει λανθασμένη ταξινόμηση, δηλαδή το μοντέλο μας προβλέπει σωστά την κλάση του σημείου δεδομένων μας, πρέπει μόνο να ενημερώσουμε την κλίση από την παράμετρο κανονικοποίησης.

$$w = w - \alpha \cdot (2\lambda w)$$

Gradient Update—No misclassification

Όταν υπάρχει λανθασμένη ταξινόμηση, για παράδειγμα όταν το μοντέλο μας κάνει λάθος στην πρόβλεψη της κλάσης του σημείου δεδομένου δίνουμε το κόστος με την παράμετρο κανονικοποίησης για να εκτελέσουμε την ενημέρωση των κλίσεων.

$$w = w + \alpha \cdot (y_i \cdot x_i - 2\lambda w)$$

Gradient Update—Misclassification

Λογιστική Παλινδρόμηση

Με προβλήματα ταξινόμησης, θέλουμε να καθορίσουμε την πιθανότητα μια παρατήρηση να ανήκει ή όχι σε μια συγκεκριμένη κλάση. Επομένως, επιθυμούμε να εκφράσουμε την πιθανότητα με μια τιμή μεταξύ του 0 και του 1. Ένας απλός αλγόριθμος ταξινόμησης που δημιουργεί τιμές αυτής της μορφής είναι ο ταξινομητής λογιστικής παλινδρόμησης.

Ας υποθέσουμε ότι έχουμε ένα απλό πρόβλημα δυαδικής ταξινόμησης, όπως αυτό που περιγράφηκε νωρίτερα στο ίδιο Κεφάλαιο. Έστω $x = x_1, \dots, x_N$ τα διανύσματα εισόδου όπου $y_i \in \{0, 1\}$. Η συνάρτηση ενεργοποίησης του LR ταξινομητή καθορίζεται από την εφαρμογή μιας σιγμοειδούς συνάρτησης πάνω στην γραμμική παλινδρόμηση ούτως ώστε να λάβουμε την τελική απόφαση ταξινόμησης. Όπως περιγράφηκε στα SVMs:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Η συνάρτηση ενεργοποίησης της LR για ένα δοσμένο διάνυσμα x ορίζεται ως εξής:

$$h_w(x) = \sigma(w^T x) = \frac{1}{1 + e^{-w^T x}}$$

Η συνάρτηση κόστους που θέλουμε να ελαχιστοποιηθεί κατά τη διάρκεια της εκπαίδευσης είναι η εξής:

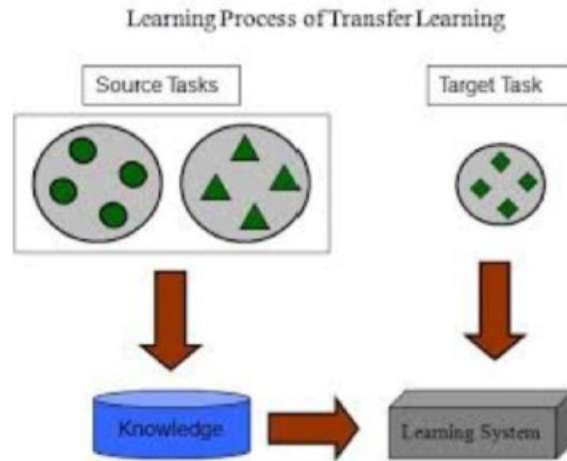
$$\frac{1}{2}\|w\| + C \sum_{i=1}^N \log(\exp(-y_i(w^T + b)) + 1)$$

όπου $C > 0$ και b είναι οι συντελεστές που αναπαριστούν την τιμωρία (penalty) των λανθασμένων αποτελεσμάτων ταξινόμησης και την τομή του υπερεπιπέδου αντίστοιχα.

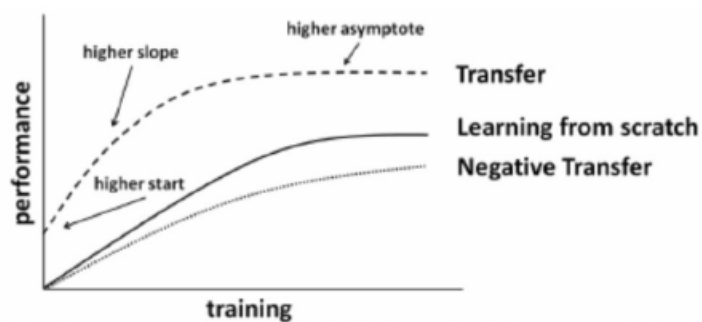
Μεταφορά Μάθησης

Μεταφορά μάθησης (Transfer Learning) ονομάζουμε το πρόβλημα της μηχανικής μάθησης, στο οποίο προσπαθούμε να αξιοποιήσουμε την γνώση που απέκτησε ένα σύστημα σε ένα πρόβλημα, σε ένα διαφορετικό αλλά σχετικό πρόβλημα. Η μέθοδος αυτή αξιοποιείται πολύ από τα βαθιά νευρωνικά δίκτυα, γιατί αυτά απαιτούν μεγάλο αριθμό δεδομένων για να εκπαιδευτούν. Θεωρητικά, άμα δεν έχουμε αρκετά δεδομένα για ένα πρόβλημα, μπορούμε να εκπαιδευσουμε ένα δίκτυο σε ένα σχετικό πρόβλημα, στο οποίο έχουμε περισσότερα δεδομένα και να χρησιμοποιήσουμε τη γνώση που απέκτησε στο αρχικό μας πρόβλημα. Στη μηχανική μάθηση (και ειδικά στη βαθιά μάθηση), αντιμετωπίζουμε ένα πολύ σημαντικό πρόβλημα. Αυτό είναι το γεγονός ότι τα δίκτυα που επιλύουν περίπλοκα προβλήματα απαιτούν τεράστιες ποσότητες δεδομένων. Ωστόσο, η απόκτηση αυτών των δεδομένων για τα επιβλεπόμενα μοντέλα είναι συχνά ανέφικτη λόγω χρονικών ή υπολογιστικών περιορισμών. Επιπλέον, τα μοντέλα που έχουν εκπαιδευτεί σε μικρά, ειδικά σύνολα δεδομένων έχουν χειρότερη απόδοση όταν χρησιμοποιούνται για να αντιμετωπίσουν ένα διαφορετικό πρόβλημα, το οποίο μπορεί να είναι σχετικά παρεμφερές με το πρόβλημα στο οποίο έχουν εκπαιδευτεί.

Ο στόχος της μεταφοράς μάθησης είναι να βελτιώσει την εκμάθηση του προβλήματος-στόχου (target task) αξιοποιώντας γνώση από το πρόβλημα-πηγή (source task), όπως φαίνεται στο παρακάτω σχήμα.



Υπάρχουν τρεις τρόποι με τους οποίους συνήθως η μεταφορά μάθησης βελτιώνει τη διαδικασία εκπαίδευσης, οι οποίοι φαίνονται στο παρακάτω. Πρώτον, η αρχική απόδοση που επιτυγχάνεται στο target task χρησιμοποιώντας μόνο τη γνώση που έχει μεταφερθεί από το source task, προτού εκπαιδευτεί παραπάνω, σε σχέση με την αρχική απόδοση ενός τυχαία αρχικοποιημένου μοντέλου. Δεύτερον, ο χρόνος που χρειάζεται για να εκπαιδευτεί πλήρως το μοντέλο στο target task δεδομένης της γνώσης που έχει μεταφερθεί, σε σχέση με το χρόνο που χρειάζεται για να το μάθει εξ αρχής. Τρίτον, το τελικό επίπεδο απόδοσης που επιτυγχάνεται στο target task σε σχέση με το τελικό επίπεδο χωρίς μεταφορά μάθησης.



Η περίπτωση της μη-επιβλεπόμενης προεκπαίδευσης (unsupervised pretraining)

Μία συγκεκριμένη περίπτωση μεταφοράς μάθησης είναι όταν το source task είναι μη επιβλεπόμενο και το target task είναι επιβλεπόμενο. Αυτή η περίπτωση έχει ιδιαίτερο ενδιαφέρον, καθώς πολύ συχνά έχουμε διαθέσιμες μεγάλες ποσότητες μη επιβλεπόμενων δεδομένων εκπαίδευσης, αλλά πολύ λίγα δεδομένα εκπαίδευσης με

ετικέτες. Η εκπαίδευση με επιβλεπόμενες τεχνικές στο επισημασμένο υποσύνολο πολλές φορές οδηγεί σε overfitting. Αποκτώντας ποιοτικές αναπαραστάσεις από τα μη επιβλεπόμενα δεδομένα, το μοντέλο μας μπορεί να έχει καλύτερη απόδοση στο πρόβλημα επιβλεπόμενης μάθησης που αντιμετωπίζουμε

Αυτή η περίπτωση μεταφοράς μάθησης ονομάζεται μη-επιβλεπόμενη προεκπαίδευση (unsupervised pretraining). Αυτή η διαδικασία αποτελεί παράδειγμα του πώς μια αναπαράσταση που έχει δημιουργηθεί από το μοντέλο, όταν αυτό αντιμετωπίζει ένα συγκεκριμένο πρόβλημα (μη επιβλεπόμενο) μπορεί κάποιες φορές να είναι χρήσιμη για ένα άλλο πρόβλημα (επιβλεπόμενο). Ονομάζεται προεκπαίδευση (pretraining), επειδή αποτελεί μόνο το πρώτο βήμα προτού ένας αλγόριθμος εκπαίδευσης εφαρμοστεί για να προσαρμόσει (fine-tune) όλα τα επίπεδα μαζί. Ως προς το πρόβλημα επιβλεπόμενης μάθησης, μπορεί να θεωρηθεί ένας όρος κανονικοποίησης και αρχικοποίησης των παραμέτρων.

Κανονικοποίηση

Είναι πιθανό ότι η προεκπαίδευση αρχικοποιεί ένα βαθύ νευρωνικό δίκτυο σε μία περιοχή που θα ήταν αλλιώς απροσπέλαστη - για παράδειγμα, μια περιοχή που περιτριγυρίζεται από περιοχές όπου η συνάρτηση κόστους εναλλάσσεται τόσο πολύ από το ένα παράδειγμα στο άλλο που μπορεί να υπολογιστεί μόνο μια εκτίμηση του gradient που περιέχει πολύ θόρυβο.

Αρχικοποίηση παραμέτρων

Η προεκπαίδευση, στις περισσότερες περιπτώσεις, βελτιώνει την απόδοση στο επιβλεπόμενο πρόβλημα. Η βασική ιδέα είναι πως κάποια χαρακτηριστικά που είναι χρήσιμα για την επίλυση του μη επιβλεπόμενου προβλήματος είναι επίσης χρήσιμα και για τη επίλυση του επιβλεπόμενου προβλήματος.

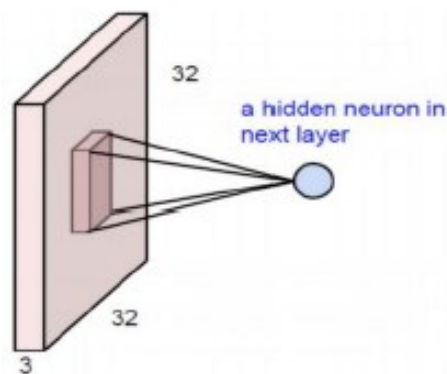
ΚΕΦΑΛΑΙΟ 10^ο

Συνελκτικά Νευρωνικά Δίκτυα: Εισαγωγή

Τα τελευταία χρόνια τα Συνελκτικά Νευρωνικά Δίκτυα(Convolutional Neural Networks-CNNs) έχουν κορυφαίες επιδόσεις στην αναγνώριση προτύπων με κυριότερη εφαρμογή τους την ταξινόμηση εικόνων. Κάτι που απασχόλησε πολύ τους ερευνητές των CNN είναι αν εντοπίζουν μόνο χαρακτηριστικά που έχουν χωρική σχέση μεταξύ τους. Μια ακόμα πολύ σημαντική πλευρά των CNN είναι ο εντοπισμός αυθαίρετων (abstract) χαρακτηριστικών καθώς οι εικόνες τροφοδοτούνται στο δίκτυο. Παρακάτω αναλύονται τα στοιχεία τους.

Στοιχεία Συνελκτικού Νευρωνικού Δικτύου

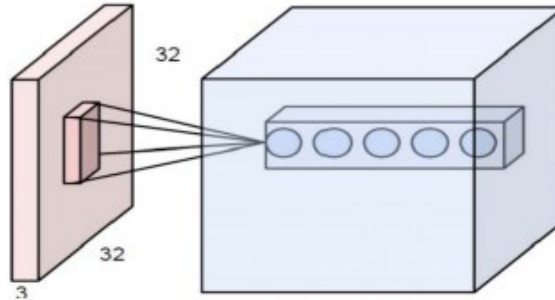
Αντί να στέλνουμε πληροφορία για ολόκληρη την εικόνα στους νευρώνες των επόμενων στοιβάδων, ψάχνουμε για μικρές περιοχές γύρω από τα εικονοστοιχεία, όπως φαίνεται στο παρακάτω σχήμα. Με άλλα λόγια οι νευρώνες των επόμενων στοιβάδων δέχονται σαν είσοδο μόνο το αντίστοιχο τμήμα εικόνας των νευρώνων της προηγούμενης τους στοιβάδας, για παράδειγμα μια 5x5 γειτονιά.



Convolution as alternative for fully connected network.

Με αυτό τον τρόπο μειώνεται πολύ το πλήθος των παραμέτρων στο μοντέλο μας. Επιπλέον, αν ρίξουμε μια ματιά πιο βαθιά, βλέπουμε ότι το σημαντικότερο πλεονέκτημα που παρέχει η συνέλιξη είναι ότι επιτρέπει τον εντοπισμό και την αναγνώριση χαρακτηριστικών ανεξαρτήτως της θέσης τους στην εικόνα. Γι αυτό το λόγο αποκαλούνται συνέλιξεις.

Για να γίνει ακόμα πιο αποδοτική η ιδέα της συνέλιξης, χρησιμοποιούμε πολλές στοιβάδες σε κάθε στρώμα του CNN, ώστε να έχουμε τη δυνατότητα να χρησιμοποιήσουμε πολλαπλά φίλτρα και άρα να εξάγουμε διαφορετικά χαρακτηριστικά από την εικόνα μας.



Multiple layers which each of them correspond to different filter but looking at the same region in the given image

Η ιδέα των πολλαπλών φίλτρων ανά στρώμα φαίνεται στη παραπάνω εικόνα.

Stride

Στην πραγματικότητα το CNN δίνει πολλές επιλογές για μείωση των παραμέτρων του μοντέλου και ταυτόχρονα για την αποφυγή παράπλευρων συνεπειών. Μια από αυτές τις επιλογές είναι το stride. Επειδή υπάρχουν πολλές επικαλύψεις (overlaps) μεταξύ των συνέλιξεων των γειτονικών εικονοστοιχείων, το stride μας δίνει τη δυνατότητα να μετακινούμε με βήμα μεγαλύτερο του ενός τον πυρήνα συνέλιξης και άρα να μην επανεξετάζουμε κάποια εικονοστοιχεία. Έτσι, δε μειώνουμε μόνο την επικάλυψη αλλά και τις διαστάσεις της επόμενης στοιβάδας.

Padding

Ένα από τα μειονεκτήματα της συνέλιξης είναι η έλλειψη πληροφορίας για την γειτονική περιοχή των εικονοστοιχείων στα άκρα(borders) της εικόνας. Η λύση σε αυτό το ζήτημα είναι η προσθήκη μηδενικών(zero padding) έξω από τα σύνορα της εικόνας ώστε μπορεί να γίνει η συνέλιξη και στα άκρα της. Συγκεκριμένα, προσθέτουμε τόσες σειρές και στήλες μηδενικών όσο το μισό της διάστασης του πυρήνα του φίλτρου, όπως φαίνεται στη παρακάτω εικόνα.

0	0	0	0	0	0	0	0	0
0								0
0								0
0								0
0								0
0								0
0								0
0								0
0	0	0	0	0	0	0	0	0

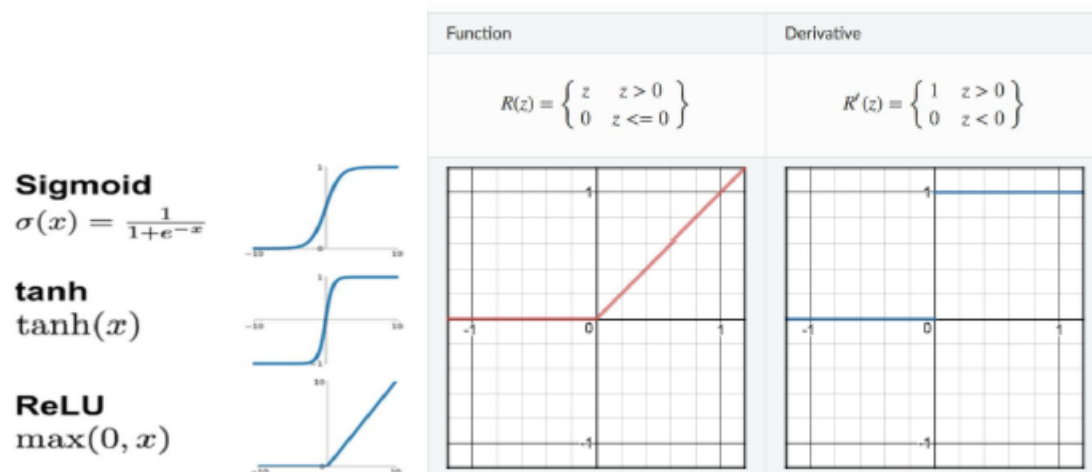
Οπότε, η διάσταση της εξόδου μιας εικόνας διάστασης W , ύστερα από P γραμμές(ή στήλες) zero padding, stride S και διάσταση φίλτρου K είναι:

$$O = \frac{(W - K + 2P)}{S} + 1$$

Μη-Γραμμικότητα

Η επόμενη στοιβάδα ύστερα από την συνελκτική είναι η μη-γραμμικότητα. Χρησιμοποιείται για να προσαρμόσει ή να κόψει (cut-off) την παραγόμενη έξοδο. Ο ρόλος της στοιβάδας αυτής είναι είτε να διαβρέξει(saturate) είτε να περιορίσει την προηγούμενη έξοδο. Για πολλά χρόνια δημοφιλής ήταν η σιγμοειδής(sigmoid) ή η υπερβολική εφαπτομένη(tanh). Ωστόσο, πρόσφατα χρησιμοποιείται η Rectified Linear Unit (ReLU) για τους εξής λόγους:

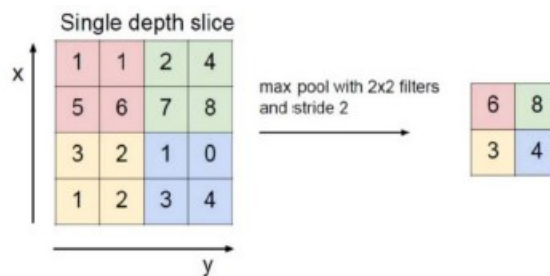
- Έχει απλούστερο ορισμό και σαν συνάρτηση και η κλίση της.



- • Η σιγμοειδής και η υπερβολική εφαπτομένη δημιουργούν το λεγόμενο vanishing gradient πρόβλημα όσο πιο βαθιά είναι η αρχιτεκτονική του δικτύου. Αυτό το αποφεύγει η ReLU καθώς έχει σταθερή θετική κλίση.
- • Η ReLU δημιουργεί μια πιο αραιή αναπαράσταση(sparse representation) καθώς η μηδενική κλίση οδηγεί σε απόκτηση ολόκληρων μηδενικών.

Στοιβάδα Pooling

Η κύρια ιδέα είναι υπο-δειγματοληψία για μείωση της πολυπλοκότητας των χαρακτηριστικών στις στοιβάδες που ακολουθούν. Στις εικόνες, αυτό σημαίνει μείωση της ανάλυσης. Η Pooling δεν επηρεάζει τον αριθμό των φίλτρων. Η Max-Pooling είναι η πιο διαδεδομένη στοιβάδα Pooling. Χωρίζει την εικόνα σε υπό-περιοχές και κρατάει μόνο τη μεγαλύτερη τιμή από την κάθε υπο-περιοχή, όπως φαίνεται στη παρακάτω εικόνα:



Πρέπει να τονιστεί ότι η υπο-δειγματοληψία δεν μειώνει την χωρική πληροφορία της εικόνας.

Πλήρως Συνδεδεμένη Στοιβάδα

Η πλήρως συνελκτική στοιβάδα(fully connected layer-FC) είναι στοιβάδα που προέρχεται από τα παραδοσιακά νευρωνικά δίκτυα. Οπότε κάθε νευρώνας στην FC στοιβάδα συνδέεται με κάθε νευρώνα της προηγούμενης και της επόμενης στοιβάδας. Το κύριο μειονέκτημα των πλήρως συνδεδεμένων στοιβάδων είναι η πολυπλοκότητα των παραμέτρων, η οποία κάνει πολύ αργή την εκπαίδευση. Οπότε μειώνουμε όσο μπορούμε το πλήθος των κόμβων της στοιβάδας αυτής με τη χρήση της τεχνικής dropout.

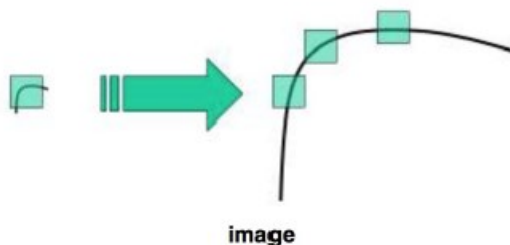
ΚΕΦΑΛΑΙΟ 11^ο

Προτεινόμενο Σύστημα

Στο κεφάλαιο αυτό αναλύουμε τις μεθόδους που χρησιμοποιήσαμε στις δύο διαφορετικές προσεγγίσεις. Αρχικά κάνουμε μια αναφορά στο σύνολο δεδομένων που χρησιμοποιήθηκε για τη διεξαγωγή των πειραμάτων. Ύστερα, προτείνουμε αρχιτεκτονικές υπολογιστικής όρασης και βαθιάς μάθησης για την όσο το δυνατόν καλύτερη προσέγγιση του προβλήματος

Αναλλοίωτος σε Κλίμακα Μετασχηματισμός

Σε προηγούμενο κεφάλαιο είδαμε τον ανιχνευτή γωνιών Harris. Είναι αμετάβλητος σε περιστροφή, που σημαίνει ότι αν μια εικόνα περιστραφεί, θα βρούμε τις ίδιες γωνίες. Είναι προφανές καθώς οι γωνίες παραμένουν γωνίες αν περιστρέψουμε την εικόνα. Αλλά σε ότι αφορά την κλίμακα, τι συμβαίνει? Μία γωνία μπορεί να μην αποτελεί γωνία αν αλλάξει η κλίμακα στην εικόνα. Στην παρακάτω εικόνα είναι εμφανές ότι η αλλαγή κλίμακας σε μία εικόνα μπορεί να επηρεάσει τις γωνίες που εντοπίζονται.



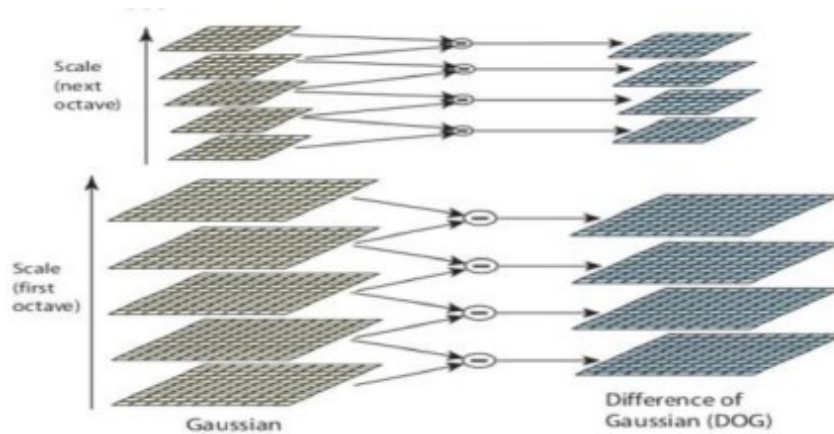
Ακολουθούμε 4 βήματα για τον αναλλοίωτο σε κλίμακα μετασχηματισμό:

1. Εύρος ανίχνευσης χώρου κλίμακας

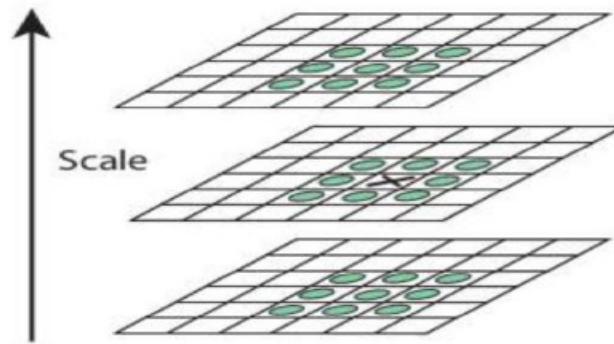
Από την παραπάνω εικόνα, είναι προφανές ότι δεν μπορούμε να χρησιμοποιήσουμε το ίδιο παράθυρο για την ανίχνευση σημείων κλειδιών με διαφορετική κλίμακα. Είναι εντάξει με μικρή γωνία. Αλλά για να ανιχνεύσουμε μεγαλύτερες γωνίες χρειαζόμαστε μεγαλύτερα παράθυρα. Για αυτό, χρησιμοποιείται φιλτράρισμα χώρου κλίμακας. Σε αυτό, βρίσκουμε το Laplacian of Gaussian για την εικόνα με διάφορες τιμές σ . Το LoG ενεργεί ως ανιχνευτής blobs, ο οποίος ανιχνεύει τα blobs σε διάφορα μεγέθη λόγω αλλαγής σ . Ητοι, το σ λειτουργεί ως παράμετρος κλιμάκωσης. Για παράδειγμα, στην

παραπάνω εικόνα, ο πυρήνας gaussian με χαμηλό σ δίνει μεγάλη αξία για τη μικρή γωνία ενώ ο πυρήνας gaussian με υψηλό σ ταιριάζει καλά για μεγαλύτερη γωνία. Έτσι, μπορούμε να βρούμε τα τοπικά μέγιστα σε όλη την κλίμακα και το διάστημα που μας δίνει μια λίστα με τιμές (x, y, σ) που σημαίνει ότι υπάρχει ένα πιθανό σημείο κλειδί στα (x, y) στην σ κλίμακα.

Αλλά το LoG είναι υπολογιστικά ακριβό, οπότε ο αλγόριθμος SIFT χρησιμοποιεί τη διαφορά Gaussians που είναι μια προσέγγιση του LoG. Η διαφορά του Gaussian λαμβάνεται ως η διαφορά Gaussian θόλωσης μιας εικόνας με δύο διαφορετικές σ , ως είναι σ και $k\sigma$. Αυτή η διαδικασία γίνεται για διαφορετικές οκτάβες της εικόνας στη Γκαουσιανή Πυραμίδα. Αναπαρίσταται στην παρακάτω εικόνα:



Μόλις εντοπιστεί το DoG, στις εικόνες αναζητούνται τοπικά άκρα σε κλίμακα και χώρο. Για παράδειγμα, ένα εικονοστοιχείο σε μια εικόνα συγκρίνεται με τους 8 γείτονές του, καθώς και 9 εικονοστοιχεία στην επόμενη κλίμακα και 9 εικονοστοιχεία σε προηγούμενες κλίμακες. Αν είναι τοπικό ακρότατο, είναι ένα πιθανό σημείο κλειδί. Βασικά αυτό σημαίνει ότι το σημείο κλειδί αντιπροσωπεύεται καλύτερα σε αυτή την κλίμακα. Αυτό φαίνεται παρακάτω:



2. Τοποθέτηση Σημείου κλειδί

Μόλις εντοπιστούν οι πιθανές τοποθεσίες σημείων κλειδιών, πρέπει να αναθεωρηθούν για να επιτευχθούν ακριβέστερα αποτελέσματα. Χρησιμοποιούμε σειρές Taylor επέκτασης κλίμακας χώρου για να πάρουμε ακριβέστερη θέση ακραίων τιμών και εάν η ένταση σε αυτά τα όρια είναι μικρότερη από μια τιμή κατωφλίου, απορρίπτεται.

Το DoG έχει υψηλότερη απόκριση για τις εκμές, επομένως και οι ακμές πρέπει να αφαιρεθούν. Για αυτό, χρησιμοποιείται μια λογική παρόμοια με την ανίχνευση γωνίας Harris. Χρησιμοποιήθηκε ένας 2×2 Hessian matrix (H) για να υπολογιστεί η κύρια καμπυλότητα. Γνωρίζουμε από τον ανιχνευτή γωνιών Harris ότι για τις άκρες, μία ιδιοτιμή είναι μεγαλύτερη από την άλλη. Έτσι, εδώ χρησιμοποιήθηκε μια απλή λειτουργία:

Αν αυτός το κλάσμα είναι μεγαλύτερο από ένα κατώφλι, τότε το σημείο κλειδί απορρίπτεται. Έτσι εξαλείφει όλα τα σημεία κλειδί χαμηλής αντίθεσης και τα σημεία κλειδί των ακμών και αυτό που παραμένει είναι τα ισχυρά σημεία ενδιαφέροντος.

3. Ανάθεση Προσανατολισμού

Τώρα ανατίθεται ένας προσανατολισμός σε κάθε βασικό σημείο για να επιτευχθεί η αμεταβλητότητα στην περιστροφή της εικόνας. Μια γειτονιά λαμβάνεται γύρω από τη θέση του βασικού σημείου ανάλογα με την κλίμακα και το μέγεθος και η κατεύθυνση της κλίσης υπολογίζονται σε αυτή την περιοχή. Παρασκευάζεται ιστόγραμμα προσανατολισμού με 36 κουτιά που καλύπτουν 360 μοίρες. Είναι σταθμισμένο από το μέγεθος κλίσης και το κυβικό παράθυρο που έχει σταθμιστεί με Gauss, με σ ίσο με 1,5 φορά την κλίμακα του βασικού σημείου. Λαμβάνεται η υψηλότερη κορυφή στο

ιστόγραμμα και θεωρείται επίσης ότι υπολογίζεται ο προσανατολισμός κάθε κορυφής άνω του 80%. Δημιουργεί σημεία κλειδιά με την ίδια θέση και κλίμακα, αλλά διαφορετικές κατευθύνσεις. Συμβάλλει στη σταθερότητα της αντιστοίχισης.

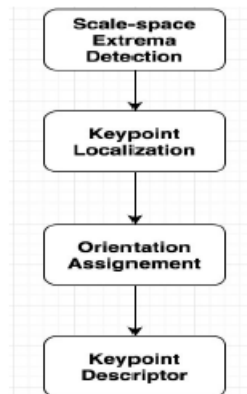
4. Περιγραφητής Σημείου κλειδι

Τώρα δημιουργείται ο περιγραφητής του σημείου κλειδι. Μια γειτονιά 16x16 γύρω από το σημείο έχει ληφθεί. Είναι χωρισμένο σε 16 υπο-τετράγωνα μεγέθους 4x4. Για κάθε υπό-ομάδα δημιουργείται ιστόγραμμα προσανατολισμού 8 κουτιών. Επομένως, είναι διαθέσιμες συνολικά 128 τιμές κουτιού και χρησιμοποιούνται ως άνυσμα για την αναπαράσταση του περιγραφητή του σημείου κλειδί. Επιπλέον, λαμβάνονται διάφορα μέτρα για την επίτευξη ευρωστίας έναντι αλλαγών φωτισμού, περιστροφής κλπ.

Η ανίχνευση των σημείων κλειδί και της κλίμακάς τους φαίνεται στην παρακάτω εικόνα:



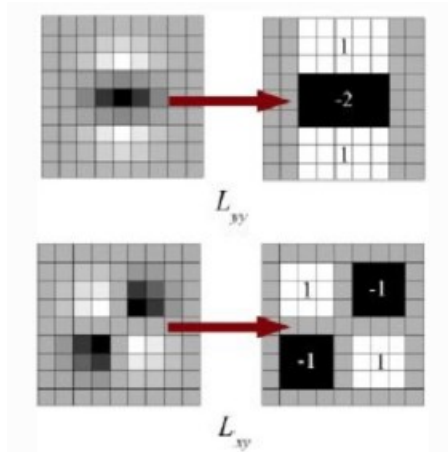
Παρακάτω φαίνεται συνοπτικά η διαδικασία που ακολουθούμε για τον υπολογισμό των SIFT χαρακτηριστικών και των περιγραφητών τους.



Επιταχυνόμενος Μετασχηματισμός

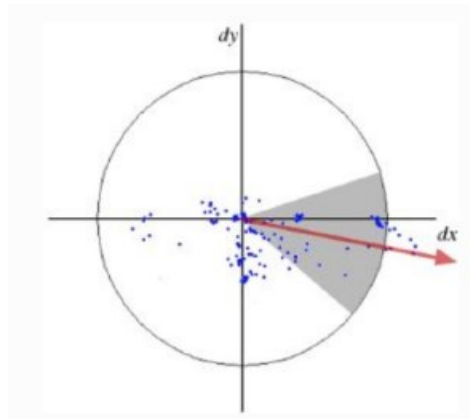
Στο κεφάλαιο αυτό, είδαμε το SIFT για την ανίχνευση και περιγραφή βασικών σημείων. Αλλά ήταν συγκριτικά αργή μέθοδος και χρειαζόμαστε πιο γρήγορη έκδοση. Το 2006, τρία άτομα, Bay, H., Tuytelaars, T. και Van Gool, L, δημοσίευσαν ένα άλλο έγγραφο, "SURF: Speed-up Robust Features"[5], που εισήγαγε ένα νέο αλγόριθμο που ονομάζεται SURF. Όπως υποδηλώνει το όνομα, πρόκειται για μια επιταχυνόμενη έκδοση του SIFT.

Στο SIFT, ο Lowe προσέγγισε τον Laplacian του Gaussian με τη διαφορά του Gaussian για την εύρεση κλίμακας χώρου. Ο SURF το διευρύνει λίγο περισσότερο και προσεγγίζει το LoG με το Φίλτρο Κουτιού. Η παρακάτω εικόνα δείχνει μια τέτοια προσέγγιση. Ένα μεγάλο πλεονέκτημα αυτής της προσέγγισης είναι ότι η συνέλιξη με το φίλτρο κουτιού μπορεί εύκολα να υπολογιστεί με τη βοήθεια ολοκληρωμένων εικόνων και μπορεί να γίνει παράλληλα για διαφορετικές κλίμακες. Επίσης, το SURF βασίζεται στον προσδιορισμό του Hessian πίνακα τόσο για την κλίμακα όσο και για την τοποθεσία.



Για ανάθεση προσανατολισμού, ο SURF χρησιμοποιεί αποκρίσεις wavelet σε οριζόντια και κάθετη κατεύθυνση. Επίσης, εφαρμόζονται επαρκή βάρη gaussian. Στη συνέχεια, σχεδιάζονται σε ένα χώρο όπως παρουσιάζεται στην παρακάτω εικόνα. Ο κυρίαρχος προσανατολισμός υπολογίζεται με τον υπολογισμό του αθροίσματος όλων των αποκρίσεων μέσα σε ένα παράθυρο ολίσθησης με γωνία 60 μοίρες. Ενδιαφέρον είναι ότι, η απόκριση wavelet μπορεί να βρεθεί χρησιμοποιώντας ολοκληρωμένες εικόνες πολύ εύκολα σε οποιαδήποτε κλίμακα. Για πολλές εφαρμογές, δεν απαιτείται μεταστροφή της περιστροφής, οπότε δεν χρειάζεται να βρεθεί αυτός ο προσανατολισμός, ο οποίος επιταχύνει τη διαδικασία. Το SURF παρέχει μια τέτοια λειτουργικότητα που ονομάζεται

Upright-SURF ή U-SURF. Βελτιώνει την ταχύτητα και είναι ανθεκτικός μέχρι και ± 15 μοίρες.

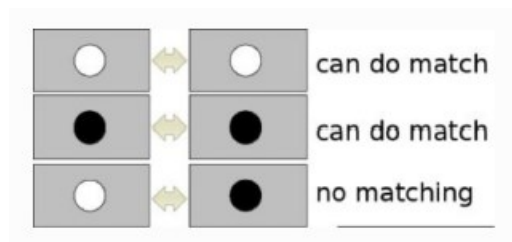


Για την περιγραφή χαρακτηριστικών, ο SURF χρησιμοποιεί αποκρίσεις Wavelet σε οριζόντια και κάθετη κατεύθυνση (και πάλι, η χρήση ολοκληρωμένων εικόνων διευκολύνει τα πράγματα). Μια γειτονιά μεγέθους 20x20 λαμβάνεται γύρω από το σημείο κλειδί όπου σ είναι η κλίμακα. Διαιρείται σε υποπεριοχές 4x4. Για κάθε υποπεριοχή, λαμβάνονται οριζόντιες και κάθετες αποκρίσεις wavelet και σχηματίζεται ένα διάνυσμα όπως αυτό:

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|).$$

Αυτό όταν αντιπροσωπεύεται ως διάνυσμα δίνει τον περιγραφητή χαρακτηριστικών SURF με συνολικά 64 διαστάσεις. Μειώστε τη διάσταση, αυξήστε την ταχύτητα του υπολογισμού και της αντιστοίχισης, αλλά παρέχετε καλύτερη διακριτικότητα των χαρακτηριστικών.

Μια σημαντική βελτίωση είναι η χρήση σημείου Laplacian (ίχνος Hessian Matrix) για το υποκείμενο σημείο ενδιαφέροντος. Δεν προσθέτει κανένα κόστος υπολογισμού, αφού έχει ήδη υπολογιστεί κατά τη διάρκεια της ανίχνευσης. Το σημάδι του Laplacian διακρίνει τις φωτεινά blobs στο σκοτεινό υπόβαθρο από την αντίστροφη κατάσταση. Στο στάδιο αντιστοίχισης, συγκρίνουμε μόνο τα χαρακτηριστικά εάν έχουν τον ίδιο τύπο αντίθεσης (όπως φαίνεται στην παρακάτω εικόνα). Αυτές οι ελάχιστες πληροφορίες επιτρέπουν την ταχύτερη αντιστοίχιση, χωρίς να μειώνεται η απόδοση του περιγραφέα.



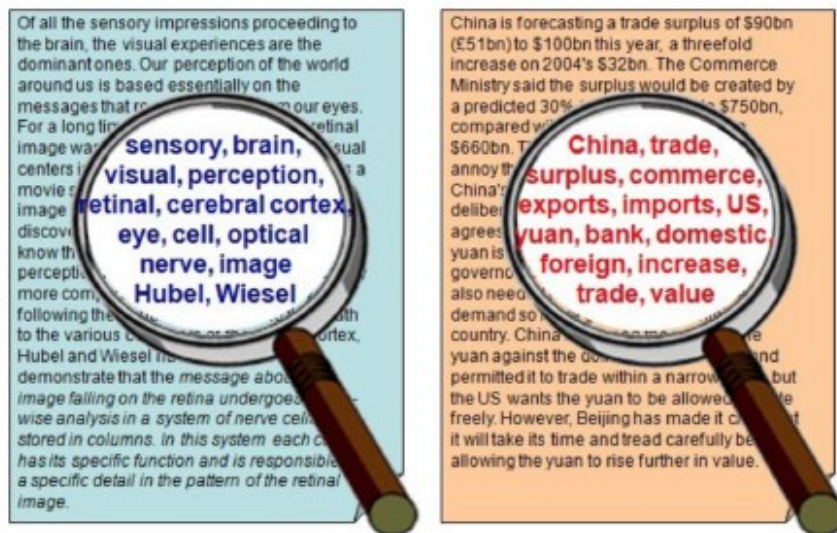
Με λίγα λόγια, ο SURF προσθέτει πολλά χαρακτηριστικά για τη βελτίωση της ταχύτητας σε κάθε βήμα. Η ανάλυση δείχνει ότι είναι 3 φορές ταχύτερη από το SIFT, ενώ η απόδοση είναι συγκρίσιμη με το SIFT. Το SURF είναι καλό στο χειρισμό εικόνων με θόλωση και περιστροφή, αλλά δεν είναι καλό στο χειρισμό της αλλαγής όψεων και της αλλαγής του φωτισμού.

Παρακάτω φαίνεται ένα παράδειγμα υπολογισμού των SURF χαρακτηριστικών με υψηλο Hessian κατωφλι (για λογους αναπαραστασης)



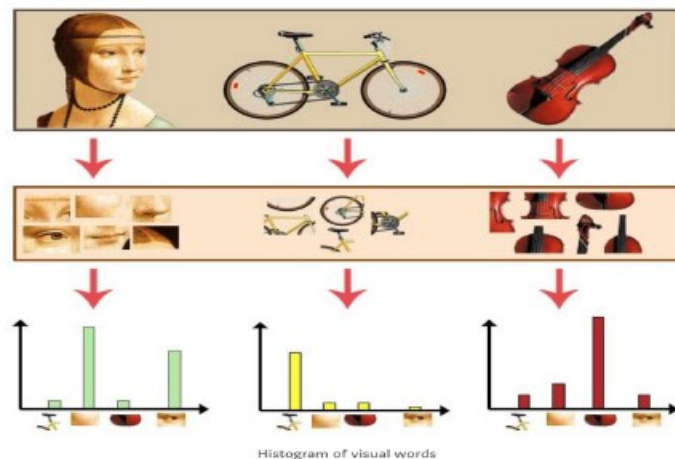
Τσάντα Οπτικών Λέξεων

Η τσάντα με οπτικές λέξεις (bag of visual words-BOVW) [8] χρησιμοποιείται συνήθως στην ταξινόμηση εικόνων. Η ιδέα της είναι υιοθετημένη από την ανάκτηση πληροφοριών και την τσάντα λέξεων της επεξεργασίας φυσικής γλώσσας. Στην τσάντα των λέξεων (bag of words-BOW) μετράμε τον αριθμό εμφανίσεων κάθε λέξης που υπάρχει σε ένα έγγραφο. Χρησιμοποιούμε τη συχνότητα κάθε λέξης για να μάθουμε τις λέξεις-κλειδιά του εγγράφου και δημιουργούμε ένα ιστόγραμμα συχνοτήτων από αυτό. Αντιμετωπίζουμε ένα έγγραφο ως BOW. Έχουμε την ίδια ιδέα στο BOVW, αλλά αντί για λέξεις, χρησιμοποιούμε χαρακτηριστικά εικόνας ως λέξεις. Τα χαρακτηριστικά εικόνας είναι μοναδικό πρότυπο που μπορούμε να βρούμε σε μια εικόνα.



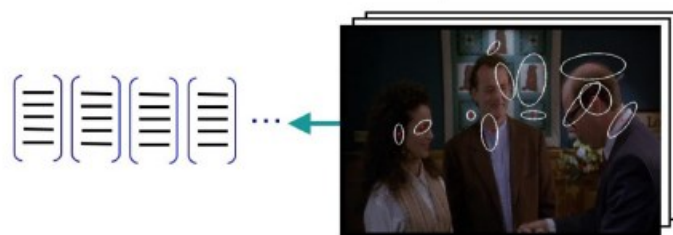
Keywords in documents

Η γενική ιδέα της τσάντας των οπτικών λέξεων (BOVW) είναι να αντιπροσωπεύει μια εικόνα ως σύνολο χαρακτηριστικών. Τα χαρακτηριστικά γνωρίσματα αποτελούνται από σημεία-κλειδιά και περιγραφητές. Τα σημεία κλειδιά είναι τα σημεία "stand-out" σε μια εικόνα, οπότε δεν έχει σημασία η εικόνα να περιστρέφεται, να συρρικνώνεται ή να επεκτείνεται, τα σημεία κλειδιά της θα είναι πάντα τα ίδια. Και ο περιγραφητής είναι η περιγραφή του σημείου-κλειδιού. Χρησιμοποιούμε τα σημεία κλειδιά και τους περιγραφητές για να κατασκευάσουμε λεξιλόγια και να αναπαριστούμε κάθε εικόνα ως ιστογράμματα συχνότητας των χαρακτηριστικών που υπάρχουν στην εικόνα. Από το ιστογράμματα συχνότητας, αργότερα, μπορούμε να βρούμε άλλες παρόμοιες εικόνες ή να προβλέψουμε την κατηγορία της εικόνας.

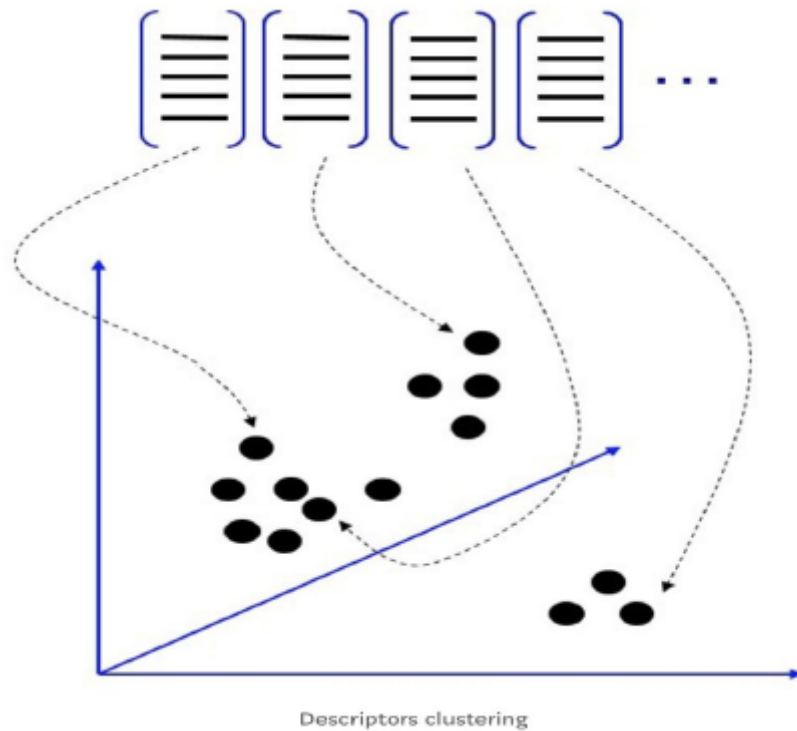


Για να χτίσουμε την τσάντα οπτικών λέξεων ακολουθούμε την εξής διαδικασία:

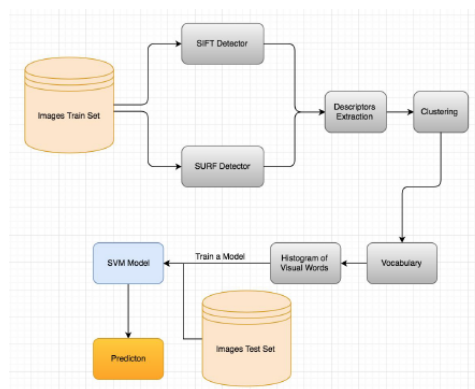
Εντοπίζουμε χαρακτηριστικά και τους περιγραφητές τους για κάθε εικόνα στο σύνολο εικόνων μας. Υστερα χτίζουμε ένα οπτικό λεξικό για κάθε εικόνα. Ο εντοπισμός των χαρακτηριστικών και ο υπολογισμός των περιγραφητών τους μπορεί να γίνει με αλγόριθμους που αναφέρθηκαν προηγουμένως (SIFT, SURF).



Ύστερα, δημιουργούμε ομάδες από τους περιγραφητές (χρησιμοποιώντας K-Means). Τα κέντρα των ομάδων μπορούν να χρησιμοποιηθούν σαν το λεξιλόγιο των οπτικών λεξικών.



Για την εκπαίδευση, χρησιμοποιούμε Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines-SVM). Ο SVM εκπαιδεύεται πάνω στα ιστογράμματα των λεξιλογίων των εικόνων προς εκπαίδευση. Για την ταξινόμηση, υπολογίζουμε αυτά τα κέντρα για τις εικόνες που θέλουμε να ταξινομήσουμε και τις κατηγοριοποιούμε ανάλογα με την απόσταση των κέντρων τους από τα κέντρα των εικόνων προς εκπαίδευση. Παρακάτω φαίνεται ένα διάγραμμα ροής της παραπάνω διαδικασίας:



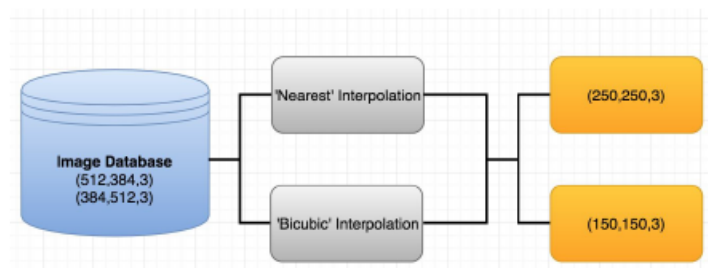
Αρχιτεκτονική Συνελικτικού Δικτύου

Τώρα θα συζητήσουμε για τις τεχνικές αναγνώρισης βασισμένες σε συνελικτικά νευρωνικά δίκτυα.

Η διαδικασία από άκρο σε άκρο έχει ως εξής.

Προεπεξεργασία εικόνας.

Σκοπός της ψηφιακής προ-επεξεργασίας εικόνας πριν αυτή εισέλθει στο νευρωνικό δίκτυο, είναι να διευκολύνει το μοντέλο να εξάγει τα χαρακτηριστικά που χρειάζεται. Χρειάζεται να εστιάσει συνήθως σε ένα μόνο συγκεκριμένο εύρος εικονοστοιχείων και η προ-επεξεργασία εικόνας βοηθάει το μοντέλο μας να επικεντρωθεί σε αυτά. Ένας ακόμα πολύ σημαντικός σκοπός της είναι η μείωση της διάστασης της εικόνας για τον περιορισμό των παραμέτρων του μοντέλου, και άρα της υπολογιστικής του πολυπλοκότητας. Παρακάτω δίνεται ένα διάγραμμα που αφορά την προ-επεξεργασία εικόνας.



Οι εικόνες της βάσης δεδομένων μας, από 512*384 συρρικνώθηκαν σε 250*250 ή ακόμα και σε 150*150.

Παρεμβολή Κοντινότερου Γείτονα

Πρόκειται για τον απλούστερο τύπο παρεμβολής. Κάθε παρεμβαλλόμενο σημείο εξόδου παίρνει τιμή από τον κοντινότερο του γείτονα στην εικόνα εισόδου. Ο πυρήνας συνέλιξης στην παρεμβολή κοντινότερου γείτονα είναι:

$$h(x) = \begin{cases} 0 & |x| > 0 \\ 1 & |x| < 0 \end{cases}$$

Η απόκριση συχνότητας του πυρήνα είναι :

$$H(\omega) = \text{sinc}(\omega/2)$$

Δικυβική Παρεμβολή

Η δικυβική παρεμβολή χρησιμοποιεί πολυώνυμα, κυβικά ή κυβικούς αλγορίθμους συνέλιξης. Η κυβική συνελκτική παρεμβολή καθορίζει την τιμή του γκριζου επιπέδου από το σταθμισμένο μέσο όρο των 16 πλησιέστερων εικονοστοιχείων στις καθορισμένες συντεταγμένες εισόδου και εκχωρεί την τιμή αυτή στις συντεταγμένες εξόδου. Για την Bicubic Interpolation, ο αριθμός των πλεγματικών σημείων που απαιτούνται για την αξιολόγηση της συνάρτησης παρεμβολής είναι 16, δύο σημεία πλέγματος σε κάθε πλευρά του σημείου κάτω τόσο για την οριζόντια όσο και για την κάθετη κατεύθυνση. Ο πυρήνας παρεμβολής δικυβικής συνέλιξης είναι:

$$W(x) = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1 & \text{for } |x| \leq 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a & \text{for } 1 < |x| < 2 \\ 0 & \text{otherwise} \end{cases}$$

όπου η παράμετρος a είναι μεταξύ -0.5 και -0.75 .

Αρχιτεκτονική CNN

Χρησιμοποιήσαμε προ-εκπαιδευμένα νευρωνικά γιατί είναι βελτιστοποιημένα και ικανά να παράγουν πολύ ισχυρά συνελκτικά χαρακτηριστικά. Συγκεκριμένα χρησιμοποιήσαμε το Inception Version 3 του GoogLeNet καθώς αυτό έδωσε τα καλύτερα αποτελέσματα.

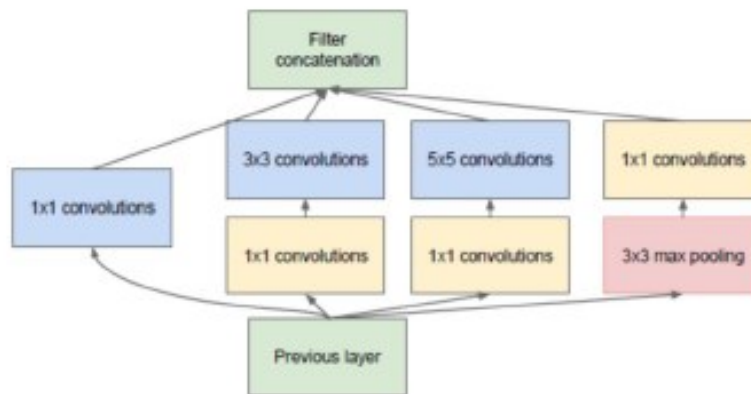
Η αρχιτεκτονική του φαίνεται παρακάτω:

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ ΓΙΑ ΤΗΝ
ΑΝΑΓΝΩΡΙΣΗ ΒΙΟΜΗΧΑΝΙΚΩΝ ΠΡΟΪΟΝΤΩΝ

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Details about Parameters of Each Layer in GoogLeNet Network (From Top to Bottom)

Περιλαμβάνει 22 στιβάδες, εκ' των οποίων εκείνη που κάνει τη διαφορά συγκριτικά με τις αρχιτεκτονικές άλλων προ-εκπαιδευμένων μοντέλων είναι η εξής:



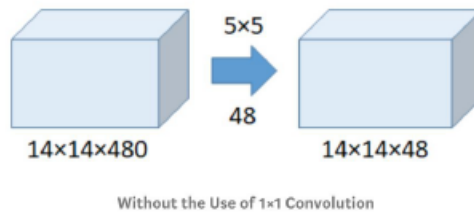
(b) Inception module with dimensionality reduction

Όπως μπορούμε να δούμε, στην είσοδο εφαρμόζονται παράλληλα

- 1X1 Συνελίξεις
- 3X3 Συνελίξεις
- 5X5 Συνελίξεις

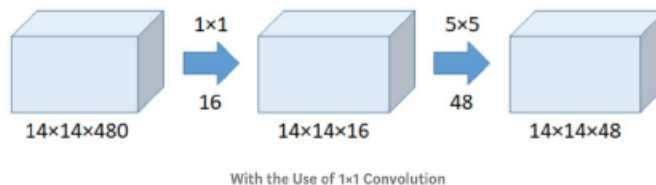
- 3X3 Max-pooling

όλες σε συνδυασμό με 1X1 συνελίξεις. Αυτό συμβαίνει γιατί οι 1X1 συνελίξεις μειώνουν την υπολογιστική πολυπλοκότητα, πετυχαίνοντας παράλληλα μη-γραμμική μείωση της διάστασης, που με τη σειρά της οδηγεί σε αποφυγή του overfitting.



$$\text{Number of operations} = (14 \times 14 \times 48) \times (5 \times 5 \times 480) = 112.9\text{M}$$

With the use of 1x1 convolution:



$$\text{Number of operations for } 1 \times 1 = (14 \times 14 \times 16) \times (1 \times 1 \times 480) = 1.5\text{M}$$

$$\text{Number of operations for } 5 \times 5 = (14 \times 14 \times 48) \times (5 \times 5 \times 16) = 3.8\text{M}$$

$$\text{Total number of operations} = 1.5\text{M} + 3.8\text{M} = 5.3\text{M which is much much smaller than } 112.9\text{M} !!!!!!!!!!!!!!!$$

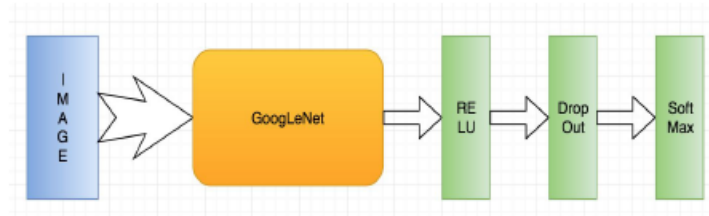
Όπως παρατηρούμε, η εφαρμογή 1X1 συνελίξης πριν από 5X5 ή οποιαδήποτε άλλη συνελίξη, βελτιώνει σημαντικά την υπολογιστική πολυπλοκότητα του μοντέλου μας. Ύστερα, οι συνελικτικοί χάρτες που προκύπτουν από τις παράλληλες συνελίξεις συνενώνονται, δίνοντας μας 4 διαφορετικά είδη χαρακτηριστικών για την είσοδο.

Στο τέλος εμείς προσθέσαμε 3 ακόμα στιβάδες:

1. ReLu (1 X 1 X 50)
2. Dropout 20%
3. Softmax (1 X 1 X Num_of_Classes)

Προσθέσαμε την στιβάδα ReLu για την ομαλότερη μετάβαση (από 1000 νευρώνες σε 50) στην στιβάδα εξόδου.

Παρακάτω φαίνεται η τελική αρχιτεκτονική του δικτύου που υλοποιήσαμε:



ΚΕΦΑΛΑΙΟ 12^ο

Διεξαγωγή Πειραμάτων

Μετρικές

Στα πειράματα μας εξάγουμε τον πίνακα σύγχυσης (confusion matrix). Ο Confusion Matrix μας δίνει πληροφορία για τις επιδόσεις της κάθε κλάσης όχι μόνο σε σχέση με την ίδια, αλλά και σε σχέση με τις υπόλοιπες κλάσεις των δεδομένων που χρησιμοποιήθηκαν για την αξιολόγηση.

Από τον confusion matrix, μπορούμε λοιπόν να εξάγουμε τις εξής τρεις μετρικές:

- Accuracy
- Precision
- Recall (or Sensitivity)

Accuracy

Η accuracy στα προβλήματα ταξινόμησης είναι το ποσοστό των σωστών προβλέψεων που έκανε το μοντέλο μας πάνω σε όλα τα είδη των προβλέψεων και ισούται με:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Στον αριθμητή έχουμε τις σωστές προβλέψεις (True Positives & True Negatives) και στον παρονομαστή έχουμε όλες τις προβλέψεις (True Positives & False Positives & False Negatives & True Negatives). Η accuracy είναι μετρική που χρησιμοποιούμε για την αξιολόγηση των πειραμάτων μας καθώς οι κλάσεις είναι ισοδύναμες.

Precision

Η precision είναι το ποσοστό των προβλέψεων που ήταν σωστές και ισούται με:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (Sensitivity)

Η recall (or sensitivity) είναι μια μετρική που δίνει το κλάσμα των σωστών προβλέψεων δια των λαθών άλλης κλάσης και ισούται με:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Πειραματικές Ρυθμίσεις

Χρησιμοποιήσαμε μηχανήμα με 2,6 GHz Intel Core i5 6 CPU's και 8 GB 1600 MHz DDR3 μνήμη. Η GPU που χρησιμοποιήσαμε ήταν η Intel Iris 1536 MB και ο τύπος του λειτουργικού συστήματος ήταν MacOS High Sierra. Για τα πειράματά μας χρησιμοποιήσαμε τα εξής frameworks:

- Tensorflow/Keras
- Sklearn,

Κατά προσέγγιση, κάθε επανάληψη στα πειράματά μας στο μηχανήμα αυτό διαρκούσε μία ώρα.

Στα πειράματά μας χρησιμοποιήσαμε 3 μόνο κλάσεις. Χωρίσαμε το σύνολο δεδομένων ως εξής:

1. Σύνολο δεδομένων προς εκπαίδευση(train set) (65%)
2. Σύνολο δεδομένων προς εξακρίβωση(validation set)(10%)
3. Σύνολο δεδομένων προς αξιολόγηση(test set)(25%)

Αποτελέσματα πειραμάτων

Μέθοδος Συνελικτικών Δικτύων

Παρακάτω φαίνεται ο πίνακας πειραμάτων/ρυθμίσεων για την αρχιτεκτονική που χρησιμοποιήσαμε βασισμένη στα συνελικτικά νευρωνικά δίκτυα:

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ ΓΙΑ ΤΗΝ
ΑΝΑΓΝΩΡΙΣΗ ΒΙΟΜΗΧΑΝΙΚΩΝ ΠΡΟΪΟΝΤΩΝ

Exp/Conf	Image Preprocess	Epochs	Validation Accuracy	Test Accuracy	Optimizer	Notes
exp01	-->(200,200) (nearest)	7	90%	88%	SGD	
exp02	-->(250,250) (nearest)	4	88%	72%	Adam	
exp03	-->(200,200) (bicubic)	7	86%	73%	SGD	
exp04	-->(150,150) (nearest)	7	88%	89%	RMSprop	Big Batch Size
exp05	-->(150,150) (nearest)	7	91%	91%	RMSProp	Small Batch Size
exp06	-->(250,250) (nearest)	3	96%	95%	RMSProp	Small Batch Size
exp07	-->(150,150) (nearest)	6	87%	88%	RMSprop	Less parameters

Όπως παρατηρούμε, ο RMSprop optimizer με παρεμβολή σε εικόνες 250x250 έδωσε ακρίβεια στο Test Set 95%. Μια ακόμα σημαντική παρατήρηση είναι ότι ο RMSprop optimizer βοήθησε στη γενίκευση του μοντέλου μας καθώς παρατηρούμε μικρές τετριμμένες διαφορές ανάμεσα στο Validation και στο Test accuracy.

ΚΕΦΑΛΑΙΟ 13^ο

Συμπεράσματα και μελλοντική εργασία

Στην παραπάνω μελέτη εξετάστηκαν δύο κύριες συνιστώσες για την προσέγγιση του προβλήματος της αισθητικής των ιστοσελίδων μέσω βαθιάς μάθησης και από μια διαφορετική σκοπιά, αυτή του δημογραφικού υποβάθρου. Η πρώτη προσέγγιση εξετάζει τη σημασία και τη δύναμη των αλγορίθμων βαθιάς μάθησης και συγκεκριμένων αρχιτεκτονικών Συνελκτικών Νευρωνικών Δικτύων, με τους οποίους γίνεται να επιτευχθούν υψηλής ποιότητας προβλέψεις της αισθητικής. Η δεύτερη προσέγγιση επιχειρεί να αξιοποιήσει τη γνώση των δημογραφικών μεταβλητών με στόχο την καλύτερη κατανόηση της επίδρασης που έχουν στη διαμόρφωση της αντίληψης της αισθητικής. Συγχρόνως, οι δημογραφικές μεταβλητές χρησιμοποιούνται για τη βελτίωση των μοντέλων βαθιάς μάθησης.

Στην πρώτη προσέγγιση, εξετάστηκαν τρεις διαφορετικές αρχιτεκτονικές ΣΝΔ πάνω στο πρόβλημα της αισθητικής των ιστοσελίδων. Η μία αρχιτεκτονική που χρησιμοποιείται ως βάση όλων των συγκρίσεων είναι επιλεγμένη από τη βιβλιογραφία. Πέρα όμως από αυτήν, χρησιμοποιούνται δύο σύγχρονες αρχιτεκτονικές ΣΝΔ οι οποίες εφαρμόζονται πρώτη φορά ως λύση στο συγκεκριμένο πρόβλημα και φαίνεται πως μπορούν να προσφέρουν βελτιωμένες προβλέψεις και καλύτερη κατανόηση των χαρακτηριστικών που επηρεάζουν την αισθητική αντίληψη του ατόμου. Οι νέες καινοτομίες που έχουν επιτευχθεί στα δίκτυα αυτά, συνεισφέρουν σε καλύτερες προβλέψεις και μέσα από τα εξεταζόμενα παραδείγματα παρατηρείται ότι τα μοντέλα συλλαμβάνουν τις συσχετίσεις μεταξύ των εικόνων και για αυτό το λόγο επιτυγχάνονται υψηλές θετικές τιμές συσχετίσεων Pearson. Το γεγονός αυτό δείχνει την αποτελεσματικότητα της βαθιάς μάθησης για την επίλυση του συγκεκριμένου προβλήματος.

Στη δεύτερη προσέγγιση το επίκεντρο της μελέτης γίνεται ο δημογραφικός παράγοντας. Αφού έχει εξετασθεί η σημαντικότητα των δημογραφικών μεταβλητών μέσω διάφορων μεθόδων, στη συνέχεια γίνεται διαχωρισμός του συνόλου δεδομένων σε υποσύνολα με βάση τις δημογραφικές μεταβλητές. Στόχος του διαχωρισμού είναι η καλύτερη μελέτη συγκεκριμένων δημογραφικών ομάδων χρησιμοποιώντας και σε αυτή την περίπτωση αλγορίθμους βαθιάς μάθησης. Εκπαιδεύοντας ένα ΣΝΔ σε κάθε

δημογραφική ομάδα, βοηθάει στην καλύτερη αντίληψη ορισμένων χαρακτηριστικών που αντιπροσωπεύουν την εκάστοτε ομάδα, ωστόσο η προβλεπτική τους ικανότητα παρουσιάζεται ελαφρώς χειρότερη. Σε αυτό το σημείο γίνεται χρήση των συνδυαστικών μεθόδων, οι οποίες βοηθούν τόσο στο να διατηρηθούν τα χαρακτηριστικά της κάθε ομάδας ενώ παράλληλα βελτιώνουν την προβλεπτική ικανότητα των μοντέλων. Το συμπέρασμα που αναδύεται από τη δεύτερη προσέγγιση είναι αφενός η τεράστια σημασία του δημογραφικού παράγοντα για την αντιμετώπιση υποκειμενικών ζητημάτων όπως είναι η αισθητική και αφετέρου ενισχύεται η Χρήση Τεχνικών βαθιάς μάθησης.

Οι αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων είναι τόσο ισχυρές για την αναπαράσταση μιας εικόνας σαν ένα σύνολο χαρακτηριστικών που οι μη-βασιζόμενες σε αυτά μέθοδοι υστερούν πολύ σε ακρίβεια. Αυτό συμβαίνει γιατί τα CNNs έχουν τη δυνατότητα να χρησιμοποιήσουν πολλών ειδών φίλτρα και να παράγουν χάρτες χαρακτηριστικών (feature maps) οι οποίοι αναπαριστούν πολλά και πολύ ισχυρά χαρακτηριστικά της εικόνας.

Υπάρχουν αρκετές μελλοντικές προτάσεις για βελτίωση του συστήματος μας. Αρχικά, θα μπορούσαν να δοκιμαστούν και οι άλλες προ-εκπαιδευμένες αρχιτεκτονικές όπως είναι η AlexNet, η ZF Net, η ResNet και η VGG Net. Επιπλέον, μπορεί στη θέση της πλήρως συνελκτικής στοιβάδας στο τελικό επίπεδο του συνελκτικού μας δικτύου να δοκιμαστεί κάποιος ταξινομητής, όπως SVM

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Chris Harris & Mike Stephens A COMBINED CORNER AND EDGE DETECTOR, Plessey Research Roke Manor, United Kingdom © The Plessey Company pic. 19
- [2] Theodoros Evgeniou and Massimiliano Pontil , SUPPORT VECTOR MACHINES:THEORY AND APPLICATIONS ,Center for Biological and Computational Learning, and Artificial Intelligence Laboratory, MIT, E25-201, Cambridge, MA 02139, USA
- [3] Sinno Jialin Pan and Qiang Yang A Survey on Transfer Learning Fellow, IEEE
- [4] David G. Lowe Distinctive, Image Features from Scale-Invariant Keypoints, Computer Science Department University of British Columbia Vancouver, B.C., Canada lowe@cs.ubc.ca January 5, 2004
- [5] Herbert Bay Tinne Tuytelaars, and Luc Van Gool, SURF: Speeded Up Robust Features ,ETH Zurich {bay, vangool}@vision.ee.ethz.ch 2 Katholieke Universiteit Leuven {Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be
- [6] Brosnan, T., and Sun, D.-W. Improving quality inspection of food products by computer vision—a review. Journal of food engineering 61, 1 (2004), 3–16.
- [7] Brynjolfsson, E., and McAfee, A. The big data boom is the innovation story of our time. The Atlantic 21 (2011).
- [8] David Aldavert, Marçal Rusinol, Ricardo Toledo, Josep Lladós, A Study of Bag-of-Visual-Words Representations for Handwritten Keyword Spotting
- [9] Dianyuan Han, Comparison of Commonly Used Image Interpolation Methods, Dept. of Computer Engineering Wei Fang University Shandong 261061, China wfhd@163.com
- [10] Cisco. Cisco visual networking index: Forecast and methodology, 2016

to 2021 (white paper), 2017.

[11] Piji Li, Optimization Algorithms for Deep Learning, Optimization Algorithms for Deep Learning Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong pjli@se.cuhk.edu.hk

[12] Khoso, M. How much data is produced every day?<http://www.northeastern.edu/levelblog/2016/05/13/>

how-much-data-produced-every-day, 2016. [Online; accessed 01.06.2018].

[13] Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., and Soares, J. V. Leafsnap: A computer vision system for automatic plant species identification. In Computer vision–ECCV 2012. Springer, 2012, pp. 502–516.

[14] UN. Tenfold increase in childhood and adolescent obesity in four decades: new study by imperial college london and who.

<http://www.who.int/en/news-room/detail/11-10-2017-tenfold-increase-in-childhood-and-adolescent-obesity-in-four-decades-n> 2017. [Online; accessed 01.06.2018].

[15] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[16] Zheng, L., Yang, Y., and Tian, Q. Sift meets cnn: A decade survey of instance retrieval. IEEE transactions on pattern analysis and machine intelligence (2017).

[17] Zeiler, M. D., and Fergus, R. Visualizing and understanding convolutional networks. In European conference on computer vision (2014), Springer, pp. 818–833.

[18] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).

[19] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., et al. Going deeper with convolutions. Cvpr.

- [20] He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (2015), pp. 1026–1034.
- [21] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115, 3 (2015),211–252.
- [22] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12(Jul):2121–2159, 2011.
- [23] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.
- [24] Lisa Torrey and Jude Shavlik. «Transfer learning». In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI Global, 2010, pp. 242–264.
- [25] Matthew D Zeiler. Adadelata: an adaptive learning rate method. arXiv preprint arXiv:1212.5701, 2012.
- [26] Geoffrey Hinton, N Srivastava, and Kevin Swersky. Lecture 6a overview of mini–batch gradient descent. Coursera Lecture slides <https://class.coursera.org/neuralnets-2012-001/lecture>,[Online, 2012.
- [27] Piji Li, Optimization Algorithms for Deep Learning, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong pjli@se.cuhk.edu.hk
- [28] Saad ALBAWI , Tareq Abed MOHAMMED, Understanding of a Convolutional Neural Network , Department of Computer Engineering Faculty of Engineering and Architecture Istanbul Kemerburgaz University Istanbul, Turkey Saad AL-ZAWI Department of Electronic Engineering Faculty of Engineering Diyala University Diyala , Iraq , ICET2017, Antalya, Turkey 978-1-5386-1949-0/17/\$31.00 ©2017 IEEE

- [29] I. Kokkinos, E. C. Paris, and G. Group, “Introduction to Deep Learning Convolutional Networks, Dropout, Maxout 1,” pp. 1–70.
- [30] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097- 1105).
- [31] D. Stutz and L. Beyer, “Understanding Convolutional Neural Networks,” 2014.