



ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ
ΕΝΙΧΝΕΥΣΗΣ ΚΑΙ
ΕΝΤΟΠΙΣΜΟΥ
ΑΝΤΙΚΕΙΜΕΝΩΝ ΜΕ ΤΗ
ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΩΝ
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

CAPSTONE PROJECT



MARCH 12, 2024

ΛΑΜΠΟΣ ΕΥΑΓΓΕΛΟΣ-INF.DIG2102

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ: Μάρκος Αυλωνίτης, Ιωάννης Καρύδης, Αλέξανδρος Πανάρετος

ΠΕΡΙΕΧΟΜΕΝΟ

ΑΝΑΛΥΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ 3-10

ΑΝΑΛΥΣΗ ΣΥΝΕΛΙΚΤΙΚΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ 10-18

YOU ONLY LOOK ONCE 18-20

ΑΝΑΛΥΣΗ ΔΙΑΦΟΡΩΝ ΑΡΧΙΤΕΚΤΟΝΙΚΩΝ ΠΡΟΤΑΣΕΩΝ ΣΤΗΝ
ΤΜΗΜΑΤΟΠΟΙΗΣΗ ΕΙΚΟΝΑΣ 21-27

ΕΥΡΗΜΑΤΑ ΠΕΙΡΑΜΑΤΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ 28-51

ΠΕΡΙΛΗΨΗ

Σκοπός της συγκεκριμένης εργασίας είναι η ανάλυση εντοπισμού μωβ μεδουσών στις ελληνικές θάλασσες. Πιο συγκεκριμένα, ελέω του αυξημένου αριθμού των μεδουσών στις ελληνικές θάλασσες τα τελευταία χρόνια καθώς και των δυσοίωνων προβλέψεων των επιστημόνων για τα επόμενα χρόνια, το ζήτημα της ασφάλειας των λουόμενων τίθεται προς επίλυση. Σε αυτήν την εργασία, αρχικά θα καταπιαστούμε με τις πιο δημοφιλείς αρχιτεκτονικές μηχανικής μάθηση για εντοπισμό αντικειμένων. Στην συνέχεια θα εξετάσουμε τα αποτελέσματα των πειραμάτων-δοκιμών που τελέστηκαν. Τέλος θα προτείνουμε τις ενδεδειγμένες λύσεις που μπορούν να εφαρμοστούν προκειμένου να μην πληγεί η βαριά βιομηχανία της χώρας μας, που δεν είναι άλλη πέραν του τουρισμού.

Μωβ Μέδουσες

Συγκεκριμένα, η μωβ μέδουσα (*Pelagia noctiluca*) εντοπίζεται σε βάθη από 0 μέχρι 1600 μέτρα και έχει μέγιστο μήκος τα 12 εκατοστά σε διάμετρο, με 4 πλοκάμια τροφής και 8 πλοκάμια (νηματοκλωστές) περιμετρικά της καμπάνας, όπου αυτό το μήκος των πλοκαμιών μπορεί να φτάσει μέχρι 10 μέτρα, όταν τα απλώνουν για να τραφούν, ενώ όταν είναι νεαρές οι μέδουσες, αυτά τα πλοκάμια δεν είναι μακρύτερα από τα πλοκάμια τροφής. Οι νηματοκύστες και τα κνιδοκύτταρα από τα οποία συντίθενται **περιέχουν νευροτοξικές ουσίες**, αυτές ακριβώς που μπορούν να προκαλέσουν βλάβη στον ανθρώπινο οργανισμό στην περίπτωση επαφής με την επιδερμίδα.

Σχετικά με την **τροφή** τους, κυρίως τρέφονται με σάλπες, διάφορα χιτωνοφόρα, πλακτονικά καρκινοειδή και αυγά ψαριών, ενώ οι μεγαλύτεροι φυσικοί θηρευτές της μέδουσας είναι η δερματοχελώνα, η καρέτα-καρέτα, ο τόνος, ο ξιφίας και το φεγγαρόψαρο. Όταν είναι νεαρή το χρώμα της είναι πορτοκαλό-καφέ και όταν ενηλικιώνεται παίρνει αυτό το έντονο πορφυρό-μωβ χρώμα.



Συμπτώματα Τσιμπήματος Από Μωβ Μέδουσα

Οι νηματοκύστες παράγουν στο ανθρώπινο δέρμα, ερύθημα, πρήξιμο, κάψιμο όπως και μερικές φορές σοβαρές δερμονεκρωτικές, καρδιο- και νευροτοξικές επιδράσεις, οι οποίες είναι ιδιαίτερα επικίνδυνες σε ευαίσθητα άτομα.

Τα συμπτώματα μετά από επαφή-τίμπημα από τη μέδουσα μπορεί να είναι: Πόνος σαν κάψιμο, πολλές φορές έντονο κοκκίνισμα του δέρματος, και σε μερικές περιπτώσεις εμφάνιση σε μέρος του δέρματος μας το αποτύπωμα της μέδουσας, ναυτία, πτώση πίεσης, ταχυκαρδία, κεφαλαλγία, εμετός, διάρροια, σπασμός των βρόγχων, δύσπνοια. Σε περίπτωση εμφάνισης συστηματικών συμπτωμάτων (σπάνια) όπως: υπόταση, βράχος φωνής, εισπνευστικός συριγμός, γενικευμένο αγγειοοίδημα - εκτεταμένο κνιδωτικό εξάνθημα, διαταραχές επιπέδου επικοινωνίας - συνείδησης, εμετός, είναι επιβεβλημένη η άμεση διακομιδή του ασθενούς στο νοσοκομείο.

Γιατί Εμφανίζονται Τώρα

Σύμφωνα με τους επιστήμονες ένας από τους βασικούς λόγους που έχει οδηγήσει στην έξαρση του πληθυσμού των μεδουσών δεν είναι άλλος από την κλιματική αλλαγή. Η αύξηση της θερμοκρασίας των υδάτων και η οργανική ρύπανση δημιουργεί τις κατάλληλες συνθήκες για την αναπαραγωγή αυτών των οργανισμών. Αυτό το φαινόμενο είναι γνωστό στη Μεσόγειο τα τελευταία 200 χρόνια. Συνήθως εμφανίζεται σε κύκλους διάρκειας 4 έως 5 ετών ανά 10-12 χρόνια παρ' όλα αυτά με τις τρέχουσες κλιματικές συνθήκες, τα κύματα εξάρσεων αναμένονται στη χώρα μας ανά τετραετία. Για τον λόγο αυτό θα πρέπει τόσο η πολιτεία όσο και ο κάθε ένας από εμάς να συνδράμει στην εξεύρεση μιας συνύπαρξης αρμονικής συνύπαρξης μεταξύ f λουομένων και μεδουσών.

Νευρωνικά Δίκτυα

Εισαγωγή

Τα νευρωνικά δίκτυα αποτελούν το αποκορύφωμα της τεχνητής νοημοσύνης και των τεχνολογιών μηχανικής μάθησης, αντλώντας έμπνευση από τις περίπλοκες λειτουργίες του ανθρώπινου εγκεφάλου. Με την ικανότητά τους να επεξεργάζονται πολύπλοκα δεδομένα και να μαθαίνουν από μοτίβα, τα νευρωνικά δίκτυα έχουν φέρει επανάσταση σε διάφορους τομείς, από την όραση υπολογιστών έως την επεξεργασία φυσικής γλώσσας. Αυτή η εργασία εμβαθύνει στις θεμελιώδεις έννοιες, τις αρχιτεκτονικές, τις μεθόδους εκπαίδευσης, τις εφαρμογές και τις μελλοντικές προοπτικές των νευρωνικών δικτύων.

Ιστορική Εξέλιξη των Νευρωνικών Δικτύων

Η απαρχή των νευρωνικών δικτύων χρονολογείται από τα μέσα του 20ού αιώνα, με πρωτοπόρες συνεισφορές από πρωτοπόρους όπως ο Warren McCulloch και ο Walter Pitts, οι οποίοι πρότειναν το πρώτο υπολογιστικό μοντέλο νευρωνικών δικτύων. Οι επακόλουθες εξελίξεις από τον Frank Rosenblatt κατέληξαν στην ανάπτυξη του perceptron, ενός νευρωνικού δικτύου ενός στρώματος που είναι ικανό για γραμμικές εργασίες ταξινόμησης.

Ωστόσο, ο ενθουσιασμός για τα νευρωνικά δίκτυα μειώθηκε κατά τις δεκαετίες του 1970 και 1980 λόγω των περιορισμών στην υπολογιστική ισχύ και της εμφάνισης εναλλακτικών προσεγγίσεων μηχανικής μάθησης. Μόλις στα τέλη του 20ού αιώνα τα νευρωνικά δίκτυα γνώρισαν μια αναζωπύρωση, η οποία τροφοδοτήθηκε από τις ανακαλύψεις στις υπολογιστικές δυνατότητες, τις αλγοριθμικές καινοτομίες και τη συσσώρευση συνόλων δεδομένων μεγάλης κλίμακας.

Η εμφάνιση της οπισθοδιάδοσης στη δεκαετία του 1980, σε συνδυασμό με την ανάπτυξη αποτελεσματικών αλγορίθμων εκπαίδευσης και συναρτήσεων ενεργοποίησης, αναζωπύρωσε το ενδιαφέρον για τα νευρωνικά δίκτυα, ανοίγοντας το δρόμο για την ευρεία υιοθέτησή τους σε διάφορους τομείς. Τον 21ο αιώνα σημειώθηκε πρωτοφανής πρόοδος στην έρευνα των νευρωνικών δικτύων, με καταλύτη τις εξελίξεις στη βαθιά μάθηση, τους παράλληλους υπολογισμούς και τα καταναμημένα συστήματα.

Σήμερα, τα νευρωνικά δίκτυα διαπερνούν σχεδόν κάθε πτυχή της σύγχρονης κοινωνίας, τροφοδοτώντας ευφυή συστήματα που διευκολύνουν την επικοινωνία, τις μεταφορές, την υγειονομική περίθαλψη και την ψυχαγωγία. Καθώς αναλογιζόμαστε την ιστορική πορεία των νευρωνικών δικτύων, αποκτούμε εικόνα για τον μετασχηματιστικό αντίκτυπο της επίμονης καινοτομίας και της διεπιστημονικής συνεργασίας.

Θεμέλια των Νευρωνικών Δικτύων

Στην καρδιά των νευρωνικών δικτύων βρίσκονται οι νευρώνες, οι θεμελιώδεις υπολογιστικές μονάδες. Αυτοί οι νευρώνες οργανώνονται σε επίπεδα, τα οποία συνήθως περιλαμβάνουν ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Οι συνδέσεις μεταξύ των νευρώνων διέπονται από βάρη και προκαταλήψεις, οι οποίες καθορίζουν την ισχύ των σημάτων και επηρεάζουν την έξοδο του δικτύου. Οι συναρτήσεις ενεργοποίησης, όπως η σιγμοειδής ή η ReLU, εισάγουν μη γραμμικότητες στο δίκτυο, επιτρέποντάς του να καταγράφει πολύπλοκες σχέσεις στα δεδομένα.

Αρχιτεκτονικά Συστατικά Των Νευρωνικών Δικτύων

Στο επίκεντρο της αρχιτεκτονικής των νευρωνικών δικτύων βρίσκονται τα συστατικά τους στοιχεία, τα οποία συλλογικά επιτρέπουν τη μοντελοποίηση πολύπλοκων σχέσεων και μοτίβων στα δεδομένα. Η κατανόηση των δομικών στοιχείων των νευρωνικών δικτύων είναι απαραίτητη για την κατανόηση της λειτουργικότητας και των δυνατοτήτων τους.

Νευρώνες: Οι νευρώνες χρησιμεύουν ως οι θεμελιώδεις μονάδες επεξεργασίας εντός των νευρωνικών δικτύων, λαμβάνοντας σήματα εισόδου, εφαρμόζοντας συναρτήσεις μετασχηματισμού και διαδίδοντας σήματα εξόδου σε επόμενα επίπεδα.

Στρώματα: Τα νευρωνικά δίκτυα αποτελούνται από πολλαπλά στρώματα, καθένα από τα οποία περιλαμβάνει μια συλλογή διασυνδεδεμένων νευρώνων. Οι συνήθεις τύποι στρωμάτων περιλαμβάνουν:

Στρώμα εισόδου: Το στρώμα εισόδου λαμβάνει ακατέργαστα δεδομένα και τα διαβιβάζει στα επόμενα στρώματα για επεξεργασία.

Κρυφά στρώματα: Τα κρυφά στρώματα εκτελούν ενδιάμεσους υπολογισμούς, εξάγοντας σχετικά χαρακτηριστικά από τα δεδομένα εισόδου μέσω μη γραμμικών μετασχηματισμών.

Στρώμα εξόδου: Το στρώμα εξόδου παράγει τις τελικές προβλέψεις ή ταξινομήσεις με βάση τις επεξεργασμένες πληροφορίες από τα προηγούμενα στρώματα.

Βάρη και προκαταλήψεις: Τα βάρη και οι προκαταλήψεις αντιπροσωπεύουν τις ρυθμιζόμενες παραμέτρους των νευρωνικών δικτύων, καθορίζοντας την ισχύ των συνδέσεων μεταξύ των νευρώνων και επηρεάζοντας τα μοτίβα ενεργοποίησής τους.

Συναρτήσεις ενεργοποίησης: Οι συναρτήσεις ενεργοποίησης εισάγουν μη γραμμικότητα στα μοντέλα νευρωνικών δικτύων, επιτρέποντάς τους να αποτυπώνουν πολύπλοκες σχέσεις και να μαθαίνουν περίπλοκα μοτίβα στα δεδομένα. Οι συνήθεις συναρτήσεις ενεργοποίησης περιλαμβάνουν τις sigmoid, tanh, ReLU και softmax.

Συναρτήσεις απώλειας: Οι συναρτήσεις απώλειών ποσοτικοποιούν τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών εξόδων, χρησιμεύοντας ως αντικειμενική συνάρτηση κατά τη διαδικασία εκπαίδευσης. Η ελαχιστοποίηση της τιμής της

συνάρτησης απωλειών καθοδηγεί τη βελτιστοποίηση των παραμέτρων του νευρωνικού δικτύου.

Ενορχηστρώνοντας αυτά τα αρχιτεκτονικά στοιχεία με συντονισμένο τρόπο, τα νευρωνικά δίκτυα παρουσιάζουν αξιοσημείωτη προσαρμοστικότητα και υπολογιστική ικανότητα, διευκολύνοντας την επίλυση ποικίλων προβλημάτων του πραγματικού κόσμου.

Εκπαίδευση και Βελτιστοποίηση

Η εκπαίδευση των νευρωνικών δικτύων συνεπάγεται την επαναληπτική βελτιστοποίηση των παραμέτρων τους για την ελαχιστοποίηση των σφαλμάτων πρόβλεψης και την ενίσχυση της απόδοσης του μοντέλου. Αυτή η ενότητα εμβαθύνει στις ιδιαιτερότητες των αλγορίθμων εκπαίδευσης, των τεχνικών βελτιστοποίησης και των μεθόδων κανονικοποίησης που χρησιμοποιούνται στην εκπαίδευση νευρωνικών δικτύων.

Οπισθοδιάδοση: Η οπισθοδιάδοση είναι ένας θεμελιώδης αλγόριθμος για την εκπαίδευση νευρωνικών δικτύων, που επιτρέπει τον υπολογισμό των κλίσεων σε σχέση με τις παραμέτρους του δικτύου. Με τη διάδοση των κλίσεων σφάλματος προς τα πίσω μέσω του δικτύου, η οπισθοδιάδοση διευκολύνει τις ενημερώσεις παραμέτρων που ελαχιστοποιούν τα σφάλματα πρόβλεψης.

Κλίση κλίσης: Η κάθοδος κλίσης είναι ένας πανταχού παρών αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την επαναληπτική ενημέρωση των παραμέτρων του νευρωνικού δικτύου. Κατεβαίνοντας κατά μήκος της κλίσης της συνάρτησης απώλειας, η κάθοδος κλίσης στοχεύει στην εύρεση των βέλτιστων τιμών παραμέτρων που ελαχιστοποιούν τα σφάλματα πρόβλεψης.

Στοχαστική κάθοδος κλίσης (SGD): Η στοχαστική κάθοδος βαθμίδας ενισχύει την αποτελεσματικότητα της βελτιστοποίησης με βάση τη βαθμίδα υπολογίζοντας ενημερώσεις παραμέτρων χρησιμοποιώντας μικρά, τυχαία δειγματοληπτικά υποσύνολα δεδομένων εκπαίδευσης. Αυτή η στοχαστική προσέγγιση επιταχύνει τη σύγκλιση και επιτρέπει την εκπαίδευση σε σύνολα δεδομένων μεγάλης κλίμακας.

Mini-Batch Gradient Descent: Η φθίνουσα κλίση μίνι παρτίδας συνδυάζει τα πλεονεκτήματα της φθίνουσας κλίσης παρτίδας και της στοχαστικής φθίνουσας κλίσης υπολογίζοντας ενημερώσεις παραμέτρων χρησιμοποιώντας μικρά υποσύνολα

δεδομένων εκπαίδευσης σταθερού μεγέθους. Αυτή η προσέγγιση επιτυγχάνει ισορροπία μεταξύ της υπολογιστικής απόδοσης και της ταχύτητας σύγκλισης.

Τεχνικές κανονικοποίησης: Οι τεχνικές κανονικοποίησης μετριάζουν την υπερπροσαρμογή στα νευρωνικά δίκτυα περιορίζοντας την πολυπλοκότητα των παραμέτρων του μοντέλου. Οι συνήθεις μέθοδοι κανονικοποίησης περιλαμβάνουν την κανονικοποίηση L1 και L2, την εγκατάλειψη και την πρόωρη διακοπή, οι οποίες προωθούν τη γενίκευση και βελτιώνουν την ευρωστία του μοντέλου.

Ρύθμιση υπερπαραμέτρων: Ο συντονισμός υπερπαραμέτρων περιλαμβάνει τη βελτιστοποίηση των ρυθμίσεων διαμόρφωσης, όπως ο ρυθμός μάθησης, το μέγεθος δέσμης και η αρχιτεκτονική του δικτύου, για τη μεγιστοποίηση της απόδοσης του νευρωνικού δικτύου. Αυτοματοποιημένες τεχνικές όπως η αναζήτηση πλέγματος και η τυχαία αναζήτηση επιταχύνουν τη διαδικασία βελτιστοποίησης των υπερπαραμέτρων.

Αξιοποιώντας αυτές τις μεθοδολογίες εκπαίδευσης και τις στρατηγικές βελτιστοποίησης, οι επαγγελματίες μπορούν να εκπαιδεύσουν αποτελεσματικά τα νευρωνικά δίκτυα για να επιτύχουν κορυφαίες επιδόσεις σε ποικίλες εργασίες και σύνολα δεδομένων.

Εφαρμογές Των Νευρωνικών Δικτύων

Η ευελιξία των νευρωνικών δικτύων επιτρέπει την εφαρμογή τους σε διάφορους τομείς. Τα συνεπυγμένα νευρωνικά δίκτυα (CNN) υπερέχουν σε εργασίες αναγνώρισης εικόνων, ανίχνευσης αντικειμένων και τμηματοποίησης σκηνών. Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) είναι ικανά στην επεξεργασία διαδοχικών δεδομένων, καθιστώντας τα ανεκτίμητα για εργασίες όπως η γλωσσική μετάφραση και η ανάλυση συναισθήματος. Τα δημιουργικά αντιφατικά δίκτυα (GAN) έχουν φέρει επανάσταση στη δημιουργία εικόνων και στη μεταφορά στυλ, παράγοντας ρεαλιστικές εικόνες που δεν διακρίνονται από τις πραγματικές.

Πέρα από τους παραδοσιακούς τομείς, τα νευρωνικά δίκτυα εισβάλλουν στην υγειονομική περίθαλψη, τα χρηματοοικονομικά και τα αυτόνομα συστήματα. Στην υγειονομική περίθαλψη, βοηθούν στη διάγνωση ασθενειών, στην ανακάλυψη φαρμάκων και σε εξατομικευμένα σχέδια θεραπείας. Στα χρηματοοικονομικά, τροφοδοτούν τις αλγοριθμικές συναλλαγές, την ανίχνευση απάτης και την αξιολόγηση κινδύνων. Τα αυτόνομα οχήματα βασίζονται σε νευρωνικά δίκτυα για την αντίληψη, τη λήψη αποφάσεων και την πλοήγηση, ανοίγοντας το δρόμο για ασφαλέστερα και αποτελεσματικότερα συστήματα μεταφορών.

Μελλοντικές Κατευθύνσεις και Προκλήσεις

Καθώς τα νευρωνικά δίκτυα συνεχίζουν να εξελίσσονται, αναδύονται διάφοροι δρόμοι έρευνας και ανάπτυξης. Οι εξελίξεις στο υλικό, όπως οι εξειδικευμένοι επιταχυντές όπως οι GPU και οι TPU, επιτρέπουν την εκπαίδευση και την ανάπτυξη ολοένα και πιο πολύπλοκων μοντέλων νευρωνικών δικτύων. Οι αρχιτεκτονικές καινοτομίες, συμπεριλαμβανομένων των μηχανισμών προσοχής και των μετασχηματιστών, ενισχύουν την ικανότητα του δικτύου να συλλαμβάνει εξαρτήσεις μεγάλης εμβέλειας και το πλαίσιο των δεδομένων.

Παρά τις εξελίξεις αυτές, οι προκλήσεις εξακολουθούν να υφίστανται. Οι δεοντολογικές ανησυχίες σχετικά με την προστασία της ιδιωτικής ζωής των δεδομένων, την προκατάληψη και τη λογοδοσία απαιτούν προσεκτική εξέταση και ρύθμιση. Η ερμηνευσιμότητα των νευρωνικών δικτύων παραμένει ένα δύσκολο έργο, εμποδίζοντας την υιοθέτησή τους σε κρίσιμους τομείς όπου η εμπιστοσύνη και η διαφάνεια είναι υψίστης σημασίας. Η αντιμετώπιση αυτών των προκλήσεων απαιτεί διεπιστημονική συνεργασία, με τη συμμετοχή ερευνητών, φορέων χάραξης πολιτικής και ενδιαφερόμενων φορέων της βιομηχανίας.

Εν κατακλείδι, τα νευρωνικά δίκτυα αποτελούν μια μετασχηματιστική τεχνολογία με βαθιές επιπτώσεις στην κοινωνία. Με την προώθηση της έρευνας, την προώθηση δεοντολογικών πρακτικών και τη συνεργασία, μπορούμε να ξεκλειδώσουμε το πλήρες δυναμικό των νευρωνικών δικτύων για την αντιμετώπιση πολύπλοκων προκλήσεων, την προώθηση της καινοτομίας και τη βελτίωση της ποιότητας ζωής των ατόμων παγκοσμίως.

Νευρωνικά Δίκτυα Συνδιαμόρφωσης(CNN)

Εισαγωγή στα Συνελκτικά Νευρωνικά Δίκτυα (CNN)

Τα συνεπτυγμένα νευρωνικά δίκτυα (CNNs) αποτελούν μια επαναστατική εξέλιξη στη μηχανική μάθηση, ιδίως στον τομέα της όρασης υπολογιστών. Αυτό το δοκίμιο ξεκινάει ένα ολοκληρωμένο ταξίδι στην αρχιτεκτονική, τη λειτουργία, τις εφαρμογές και τις πρόσφατες εξελίξεις των CNNs, αναδεικνύοντας τον καθοριστικό τους ρόλο στη μεταμόρφωση του τοπίου των εργασιών οπτικής αναγνώρισης.

Τα CNN αντιπροσωπεύουν μια εξειδικευμένη κατηγορία νευρωνικών δικτύων που έχουν δημιουργηθεί σχολαστικά για να μιμούνται την αξιοσημείωτη ικανότητα του ανθρώπινου οπτικού συστήματος να αντιλαμβάνεται και να ερμηνεύει τα οπτικά ερεθίσματα. Εμπνευσμένα από τη δομή και τη λειτουργία του οπτικού φλοιού, τα CNN παρουσιάζουν μια ιεραρχική αρχιτεκτονική βελτιστοποιημένη για την εξαγωγή και την

επεξεργασία περίπλοκων οπτικών χαρακτηριστικών από ακατέργαστα δεδομένα εισόδου.

Στον πυρήνα τους, τα CNN αξιοποιούν τα συνελικτικά στρώματα για την εξαγωγή εντοπισμένων χαρακτηριστικών, τα στρώματα συγκέντρωσης για την υποδειγματοληψία χαρτών χαρακτηριστικών και τα πλήρως συνδεδεμένα στρώματα για την ενσωμάτωση των εξαχθέντων χαρακτηριστικών και την πραγματοποίηση τελικών προβλέψεων. Μέσω μιας διαδικασίας ιεραρχικής εκμάθησης χαρακτηριστικών, τα CNN μπορούν να διακρίνουν πολύπλοκα μοτίβα και δομές σε εικόνες, επιτρέποντας εργασίες όπως η ανίχνευση αντικειμένων, η ταξινόμηση εικόνων, η σημασιολογική κατάτμηση και άλλα.

Αρχιτεκτονικά Συστατικά Των Συνεπαγωγικών Νευρωνικών Δικτύων

Συνελικτικά επίπεδα: Τα στρώματα συνέλιξης, ο ακρογωνιαίος λίθος των CNN, χρησιμοποιούν φίλτρα που μπορούν να μάθουν για να εκτελέσουν πράξεις συνέλιξης σε εικόνες εισόδου. Αυτά τα φίλτρα σαρώνουν τα δεδομένα εισόδου, εξάγοντας τοπικά χαρακτηριστικά μέσω της κατανομής βαρών και των χωρικών ιεραρχιών. Κάθε φίλτρο συλλαμβάνει συγκεκριμένα μοτίβα, όπως ακμές, υφές ή σχήματα, συμβάλλοντας στην ικανότητα του δικτύου να διακρίνει σημαντικές οπτικές πληροφορίες.

Επίπεδα συγκέντρωσης: Τα στρώματα συγκέντρωσης διευκολύνουν τη χωρική μείωση της δειγματοληψίας, μειώνοντας τις διαστάσεις των χαρτών χαρακτηριστικών, διατηρώντας παράλληλα τις βασικές πληροφορίες. Οι συνήθεις λειτουργίες συγκέντρωσης περιλαμβάνουν τη μέγιστη συγκέντρωση και τη μέση συγκέντρωση, οι οποίες ενισχύουν τη μεταφραστική αναλλοίωτη και την υπολογιστική αποδοτικότητα. Με τη μείωση της χωρικής ανάλυσης των χαρτών χαρακτηριστικών, τα στρώματα συγκέντρωσης επιτρέπουν στα CNN να εστιάζουν σε σημαντικά χαρακτηριστικά και να μετριάσουν τις επιπτώσεις της υπερπροσαρμογής.

Συναρτήσεις ενεργοποίησης: Οι συναρτήσεις ενεργοποίησης εισάγουν μη γραμμικότητα στις αρχιτεκτονικές CNN, επιτρέποντάς τους να καταγράφουν σύνθετες σχέσεις και να μαθαίνουν μη γραμμικές αντιστοιχίσεις μεταξύ δεδομένων εισόδου και εξόδου. Δημοφιλείς συναρτήσεις ενεργοποίησης, όπως η ReLU (Rectified Linear Unit), η sigmoid και η tanh, διευκολύνουν τη διάδοση κλίσης και επιταχύνουν τη σύγκλιση κατά την εκπαίδευση.

Πλήρως συνδεδεμένα επίπεδα: Τα πλήρως συνδεδεμένα στρώματα χρησιμεύουν ως τα τελικά στάδια των αρχιτεκτονικών CNN, συγκεντρώνοντας χωρικά χαρακτηριστικά που εξάγονται από τα στρώματα συνελίξεων και συγκέντρωσης για τη δημιουργία προβλέψεων ή ταξινομήσεων. Αυτά τα στρώματα χρησιμοποιούν πυκνές συνδέσεις μεταξύ των νευρώνων, επιτρέποντας στο δίκτυο να συμπεραίνει σημασιολογικές πληροφορίες υψηλού επιπέδου και να κάνει ακριβείς προβλέψεις.

Εκπαίδευση Συνελικτικών Νευρωνικών Δικτύων

Η εκπαίδευση των CNN περιλαμβάνει τη βελτιστοποίηση των παραμέτρων τους, όπως τα βάρη και οι προκαταλήψεις των φίλτρων, για την ελαχιστοποίηση των σφαλμάτων πρόβλεψης και τη βελτίωση της απόδοσης του μοντέλου. Αυτή η ενότητα διερευνά τη διαδικασία εκπαίδευσης, τις τεχνικές βελτιστοποίησης και τις μεθόδους κανονικοποίησης που χρησιμοποιούνται στην εκπαίδευση CNN.

Οπισθοδιάδοση: Η οπισθοδιάδοση χρησιμεύει ως η ραχοκοκαλιά της εκπαίδευσης CNN, διευκολύνοντας τον υπολογισμό των κλίσεων σε σχέση με τις παραμέτρους του δικτύου. Με τη διάδοση των κλίσεων σφάλματος προς τα πίσω μέσω του δικτύου, η οπισθοδιάδοση επιτρέπει ενημερώσεις παραμέτρων που ελαχιστοποιούν τα σφάλματα πρόβλεψης.

Κάθοδος κλίσης: Οι αλγόριθμοι κατάβασης κλίσης, όπως η στοχαστική κάθοδος κλίσης (SGD) και ο Adam, βελτιστοποιούν τις παραμέτρους του CNN κατεβαίνοντας κατά μήκος της κλίσης μιας συνάρτησης απωλειών. Αυτές οι τεχνικές βελτιστοποίησης ενημερώνουν επαναληπτικά τις παραμέτρους για να ελαχιστοποιήσουν τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών εξόδων, οδηγώντας το δίκτυο προς τις βέλτιστες λύσεις.

Ενίσχυση δεδομένων: Οι τεχνικές επαύξησης δεδομένων επαυξάνουν τα σύνολα δεδομένων εκπαίδευσης εφαρμόζοντας μετασχηματισμούς όπως περιστροφές, μεταφράσεις και αναστροφές στις εικόνες εισόδου. Με την εισαγωγή παραλλαγών στα δεδομένα εκπαίδευσης, η επαύξηση δεδομένων ενισχύει τη γενίκευση του μοντέλου και μειώνει την υπερβολική προσαρμογή, βελτιώνοντας την ευρωστία και την απόδοση του CNN.

Μάθηση μεταφοράς: Η εκμάθηση μεταφοράς αξιοποιεί προ-εκπαιδευμένα μοντέλα CNN, όπως τα VGG, ResNet ή Inception, που έχουν εκπαιδευτεί σε σύνολα δεδομένων

μεγάλης κλίμακας όπως το ImageNet. Με τη λεπτομερή ρύθμιση των προ-εκπαιδευμένων μοντέλων σε σύνολα δεδομένων συγκεκριμένου τομέα, η εκμάθηση μεταφοράς επιταχύνει τη σύγκλιση της εκπαίδευσης και επιτρέπει την αποτελεσματική μεταφορά γνώσης, ιδίως σε σενάρια με περιορισμένα δεδομένα εκπαίδευσης.

Εφαρμογές Των Συνεπυγμένων Νευρωνικών Δικτύων

Τα συνεπυγμένα νευρωνικά δίκτυα βρίσκουν εκτεταμένες εφαρμογές σε διάφορους τομείς, φέρνοντας επανάσταση στις βιομηχανίες και καταλύοντας τις καινοτομίες στις εργασίες οπτικής αναγνώρισης. Αυτή η ενότητα διευκρινίζει τις πολύπλευρες εφαρμογές των CNN, που κυμαίνονται από την ταξινόμηση εικόνων και την ανίχνευση αντικειμένων έως την ιατρική απεικόνιση και τα αυτόνομα οχήματα.

Ταξινόμηση εικόνων: Τα CNN υπερέχουν σε εργασίες ταξινόμησης εικόνων, κατηγοριοποιώντας με ακρίβεια τις εικόνες εισόδου σε προκαθορισμένες κλάσεις ή ετικέτες. Οι εφαρμογές περιλαμβάνουν τον εντοπισμό αντικειμένων, ζώων, ορόσημων και άλλων, με προεκτάσεις που καλύπτουν το ηλεκτρονικό εμπόριο, την επιτήρηση και τη συγκράτηση περιεχομένου.

Ανίχνευση αντικειμένων: Πλαίσια ανίχνευσης αντικειμένων που βασίζονται σε CNN, όπως το Faster R-CNN, το YOLO (You Only Look Once) και το SSD (Single Shot MultiBox Detector), επιτρέπουν την ανίχνευση και τον εντοπισμό αντικειμένων σε πραγματικό χρόνο μέσα σε εικόνες ή ροές βίντεο. Αυτά τα μοντέλα διευκολύνουν εφαρμογές όπως η αυτόνομη οδήγηση, η ρομποτική και η επαυξημένη πραγματικότητα.

Σημασιολογική τμηματοποίηση: Η σημασιολογική τμηματοποίηση περιλαμβάνει την ταξινόμηση εικόνων σε επίπεδο εικονοστοιχείου σε σημασιολογικά σημαντικές περιοχές ή αντικείμενα. Τα μοντέλα τμηματοποίησης με βάση το CNN, όπως το U-Net και το FCN (Fully Convolutional Network), επιτρέπουν την ακριβή οριοθέτηση των ορίων των αντικειμένων, διευκολύνοντας την ιατρική απεικόνιση, τον αστικό σχεδιασμό και την περιβαλλοντική παρακολούθηση.

Ιατρική απεικόνιση: Τα CNN διαδραματίζουν καθοριστικό ρόλο στην ανάλυση της ιατρικής απεικόνισης, βοηθώντας τους κλινικούς γιατρούς στη διάγνωση ασθενειών, την ανίχνευση ανωμαλιών και την πρόβλεψη των αποτελεσμάτων των ασθενών. Οι εφαρμογές περιλαμβάνουν την αυτοματοποιημένη ανίχνευση παθολογίας, την

κατάτμηση όγκων και την ανακατασκευή ιατρικών εικόνων, ενισχύοντας τη διαγνωστική ακρίβεια και τη φροντίδα των ασθενών.

Πρόσφατες Εξελίξεις Στα Συνελκτικά Νευρωνικά Δίκτυα

Τα τελευταία χρόνια έχουν σημειωθεί ραγδαίες εξελίξεις στα Συγκεχυμένα Νευρωνικά Δίκτυα, οι οποίες προήλθαν από καινοτομίες στις αρχιτεκτονικές των μοντέλων, στις τεχνικές εκπαίδευσης και στα υπολογιστικά πλαίσια. Αυτή η ενότητα εξετάζει τις αξιοσημείωτες εξελίξεις που διαμορφώνουν τη μελλοντική πορεία της έρευνας και των εφαρμογών των CNN.

Μηχανισμοί προσοχής: Οι μηχανισμοί προσοχής ενισχύουν την εστίαση των CNNs σε σχετικά χαρακτηριστικά μέσα στα δεδομένα εισόδου, επιτρέποντας την επιλεκτική εξαγωγή χαρακτηριστικών και τη δυναμική στάθμιση χαρακτηριστικών. Μοντέλα όπως ο Transformer και οι μηχανισμοί αυτοπροσοχής ενισχύουν τις αρχιτεκτονικές CNN, βελτιώνοντας τις επιδόσεις σε εργασίες όπως η απόδοση τίτλων εικόνας, η οπτική απάντηση ερωτήσεων και η μηχανική μετάφραση.

Δίκτυα κάψουλας: Τα δίκτυα Capsule Networks (CapsNets) αντιπροσωπεύουν μια εναλλακτική λύση στις παραδοσιακές αρχιτεκτονικές CNN, σχεδιασμένα για την καταγραφή ιεραρχικών χωρικών σχέσεων και την εκμάθηση χαρακτηριστικών αμετάβλητων στη θέση. Τα CapsNets αξιοποιούν δυναμικούς αλγορίθμους δρομολόγησης για τη διάδοση πληροφοριών μεταξύ των καψουλών, διευκολύνοντας την ισχυρή εξαγωγή χαρακτηριστικών και τη βελτιωμένη γενίκευση.

Γενετικά αντιθετικά δίκτυα (GAN): Τα GANs ενσωματώνουν τα CNNs σε αντιθετικά πλαίσια εκπαίδευσης, διευκολύνοντας τη δημιουργία ρεαλιστικών συνθετικών εικόνων ή βίντεο. Οι εφαρμογές περιλαμβάνουν τη μετάφραση από εικόνα σε εικόνα, τη μεταφορά στυλ και την αύξηση δεδομένων, αξιοποιώντας τις διακριτικές και γενετικές ικανότητες των CNN για τη δημιουργία οπτικού περιεχομένου υψηλής πιστότητας.

Ομαδοποιημένη μάθηση: Η ομοσπονδιακή μάθηση επιτρέπει τη συνεργατική εκπαίδευση μοντέλων σε κατανεμημένες συσκευές ή πηγές δεδομένων, διατηρώντας παράλληλα το απόρρητο και την ασφάλεια των δεδομένων. Τα πλαίσια ομοσπονδιακής μάθησης που βασίζονται σε CNN αξιοποιούν τεχνικές όπως η συνάθροιση μοντέλων, η διαφορική ιδιωτικότητα και η ασφαλής συνάθροιση για την εκπαίδευση εύρωστων μοντέλων με διατήρηση της ιδιωτικότητας σε αποκεντρωμένα σύνολα δεδομένων.

Δεοντολογικά Ζητήματα Και Προκλήσεις

Εν μέσω της ταχείας εξάπλωσης των συνεπαγωγικών νευρωνικών δικτύων, οι δεοντολογικοί προβληματισμοί και οι προκλήσεις προβάλλουν σε μεγάλο βαθμό, απαιτώντας προσεκτική σκέψη και προληπτικά μέτρα για τον μετριασμό των κινδύνων και τη διασφάλιση της κοινωνικής ευημερίας. Η παρούσα ενότητα διερευνά βασικές ηθικές εκτιμήσεις και προκλήσεις που σχετίζονται με την ανάπτυξη και τη διάδοση των CNN.

Μεροληψία και δικαιοσύνη: Τα CNN ενδέχεται να διαιωνίζουν ακούσια τις προκαταλήψεις που υπάρχουν στα δεδομένα εκπαίδευσης, οδηγώντας σε προκατειλημμένες προβλέψεις και αποτελέσματα που εισάγουν διακρίσεις. Η αντιμετώπιση της προκατάληψης και η διασφάλιση της δικαιοσύνης στα μοντέλα CNN απαιτεί προσεκτική επιμέλεια των συνόλων δεδομένων, αλγοριθμική διαφάνεια και τεχνικές εκπαίδευσης με επίγνωση της δικαιοσύνης.

Ανησυχίες σχετικά με την προστασία της ιδιωτικής ζωής: Τα CNN που εκπαιδεύονται σε ευαίσθητα ή προσωπικά δεδομένα εγείρουν ανησυχίες σχετικά με την παραβίαση της ιδιωτικής ζωής και την κατάχρηση δεδομένων. Η υιοθέτηση τεχνικών διατήρησης της ιδιωτικότητας, όπως η ομοσπονδιακή μάθηση, η ασφαλής συνάθροιση και η διαφορική ιδιωτικότητα, μπορεί να μετριάσει τους κινδύνους για την ιδιωτικότητα και να υποστηρίξει τα δικαιώματα των δεδομένων των ατόμων.

Ιστορική Αναδρομή Στην Εξέλιξη Των CNN

Οι αλγόριθμοι μηχανικής μάθησης (ML) είναι γνωστό ότι μαθαίνουν την υποκείμενη σχέση στα δεδομένα και έτσι λαμβάνουν αποφάσεις χωρίς να απαιτούν ρητές οδηγίες. Στη βιβλιογραφία, διάφορες συναρπαστικές εργασίες έχουν αναφερθεί για την κατανόηση και/ή την εξομοίωση των ανθρώπινων αισθητηριακών αντιδράσεων, όπως η ομιλία και όραση (Hubel και Wiesel 1962, 1968- Ojala κ.ά. 1996- Chappelle 1998- Lowe 1999- Dalal and Triggs 2004- Bay et al. 2008- Heikkilä et al. 2009). Το 1989, μια νέα κατηγορία νευρωνικών δικτύων (NN), που ονομάστηκε Νευρωνικό Δίκτυο Συνελκτικού τύπου (CNN) (LeCun et al. 1989), η οποία έδειξε τεράστιες δυνατότητες σε εργασίες που σχετίζονται με τη μηχανική όραση (ΜΟ). Τα CNN είναι ένας από τους καλύτερους αλγορίθμους μάθησης για την κατανόηση του περιεχομένου της εικόνας και έχουν επιδείξει υποδειγματικές επιδόσεις στην κατάτμηση, ταξινόμηση, ανίχνευση και ανάκτηση εικόνων. που σχετίζονται με εργασίες. Η επιτυχία των CNN έχει τραβήξει την προσοχή πέρα από τον ακαδημαϊκό χώρο. Στη βιομηχανία, εταιρείες όπως η Google, η Microsoft, η AT&T, η NEC και οι Facebook έχουν αναπτύξει ενεργές ερευνητικές ομάδες για τη διερεύνηση νέων αρχιτεκτονικών CNN (Denget al. 2013).

Επί του παρόντος, οι περισσότεροι από τους πρωτοπόρους της επεξεργασίας εικόνας και της όρασης υπολογιστών (CV) χρησιμοποιούν βαθιά μοντέλα βασισμένα σε CNN. Το ελκυστικό χαρακτηριστικό του CNN είναι η ικανότητά του να εκμεταλλεύεται τη χωρική ή χρονική συσχέτιση σε δεδομένα. Η τοπολογία του CNN χωρίζεται σε πολλαπλά στάδια μάθησης που αποτελούνται από ένα συνδυασμό των επιπέδων συνέλιξης, των μονάδων μη γραμμικής επεξεργασίας και των επιπέδων υποδειγματοληψίας (Jarrett et al.2009). Το CNN είναι ένα προτροφοδοτούμενο πολυεπίπεδο ιεραρχικό δίκτυο, όπου κάθε επίπεδο, χρησιμοποιώντας μια τράπεζα από πυρήνες συνελικτικής ανάλυσης, εκτελεί πολλαπλούς μετασχηματισμούς (LeCun et al. 2010). Συνελικτική λειτουργία βοηθά στην εξαγωγή χρήσιμων χαρακτηριστικών από τοπικά συσχετιζόμενα σημεία δεδομένων. Η έξοδος των πυρήνων συνέλιξης ανατίθεται στη συνέχεια στη μονάδα μη γραμμικής επεξεργασίας (ενεργοποίηση συνάρτησης), η οποία όχι μόνο βοηθά στην εκμάθηση αφαιρέσεων αλλά και ενσωματώνει τη μη γραμμικότητα στο χώρο χαρακτηριστικών. Αυτή η μη γραμμικότητα δημιουργεί διαφορετικά μοτίβα ενεργοποίησης για διαφορετικές αποκρίσεις και έτσι διευκολύνει την εκμάθηση των σημασιολογικών διαφορών στις εικόνες. Η έξοδος της μη γραμμικής συνάρτησης ενεργοποίησης συνήθως ακολουθείται από υποδειγματοληψία, η οποία βοηθά στη σύνοψη των αποτελεσμάτων και καθιστά επίσης την είσοδο αναλλοίωτη σε γεωμετρικές παραμορφώσεις (Scherer et al. 2010- LeCun et al.2010). Το CNN, με την ικανότητα αυτόματης εξαγωγής χαρακτηριστικών, μειώνει την ανάγκη για ξεχωριστό χαρακτηριστικό εξαγωγής χαρακτηριστικών (Najafabadi et al. 2015). Έτσι, το CNN χωρίς εξαντλητική επεξεργασία μπορεί να μάθει καλά εσωτερική αναπαράσταση από τα ακατέργαστα εικονοστοιχεία. Αξιοσημείωτα χαρακτηριστικά του CNN είναι η ιεραρχική μάθηση, η αυτόματη εξαγωγή χαρακτηριστικών, η πολυεργασία και ο διαμοιρασμός βαρών

Το CNN ήρθε για πρώτη φορά στο προσκήνιο μέσω της εργασίας του LeCun το 1989 για την επεξεργασία τοπολογικών δεδομένων τύπου πλέγματος (εικόνες και δεδομένα χρονοσειρών) (LeCun et al. 1989; Ian Goodfellow et al.2017). Ο αρχιτεκτονικός σχεδιασμός του CNN εμπνεύστηκε από το έργο των Hubel και Wiesel και έτσι ακολουθεί ως επί το πλείστον τη βασική δομή του οπτικού φλοιού των πρωτεύοντων (Hubel και Wiesel 1962, 1968). Τα διάφορα στάδια της διαδικασίας μάθησης στο CNN παρουσιάζουν αρκετή ομοιότητα με τον πρωτεύοντα κοιλιακή οδό του οπτικού φλοιού (V1-V2-V3-V4-IT/VTC) των πρωτεύοντων (Laskar et al. 2018). Ο οπτικός φλοιός των πρωτεύοντων λαμβάνει πρώτα εισροή από την αμφιβληστροειδοτοπική περιοχή. Με τον τρόπο αυτό, το πλευρικό γεννητικό πυρήνας εκτελεί φιλτράρισμα υψηλής διέλευσης πολλαπλών κλιμάκων και ομαλοποίηση της αντίθεσης. Μετά από αυτό, η ανίχνευση πραγματοποιείται από διάφορες περιοχές του οπτικού φλοιού που κατηγοριοποιούνται ως V1, V2, V3 και V4. Στην πραγματικότητα, οι περιοχές V1 και V2 του οπτικού φλοιού είναι παρόμοιες με τα επίπεδα συνελικτικής και υποδειγματοληψίας. Στο αντιθέτως, η κατώτερη κροταφική περιοχή μοιάζει με τα ανώτερα στρώματα του CNN, γεγονός που καθιστά ένα συμπέρασμα σχετικά με την εικόνα (Grill-Spector et al. 2018).

Κατά τη διάρκεια της εκπαίδευσης, το CNN μαθαίνει μέσω του αλγορίθμου backpropagation, ρυθμίζοντας τις αλλαγές των βαρών σύμφωνα με τον στόχο. Η βελτιστοποίηση μιας αντικειμενικής συνάρτησης με τη χρήση ενός αλγορίθμου οπισθοδιάδοσης είναι παρόμοια με τη μάθηση βάσει απόκρισης του ανθρώπινου εγκεφάλου. Το πολυεπίπεδο, ιεραρχική δομή του βαθύ CNN, του δίνει τη δυνατότητα να εξάγει χαμηλές, μεσαίες και χαρακτηριστικά υψηλού επιπέδου. Τα χαρακτηριστικά υψηλού επιπέδου (πιο αφηρημένα χαρακτηριστικά) είναι ένας συνδυασμός χαμηλότερων και χαρακτηριστικών μεσαίου επιπέδου. Η ιεραρχική ικανότητα εξαγωγής χαρακτηριστικών του CNN μιμείται τη βαθιά και πολυεπίπεδη διαδικασία μάθησης του Νεοφλοιού στον ανθρώπινο εγκέφαλο, ο οποίος μαθαίνει δυναμικά χαρακτηριστικά από τα ακατέργαστα δεδομένα (Bengio 2009). Η δημοτικότητα του CNN οφείλεται κυρίως στην ιεραρχική του ικανότητα εξαγωγής χαρακτηριστικών.

Οι βαθιές αρχιτεκτονικές συχνά πλεονεκτούν έναντι των ρηχών αρχιτεκτονικών όταν έχουν να κάνουν με σύνθετα προβλήματα μάθησης. Η στοίβαξη πολλαπλών γραμμικών και μη γραμμικών μονάδων επεξεργασίας σε ένα τρόπο κατά στρώματα παρέχει τη δυνατότητα εκμάθησης σύνθετων αναπαραστάσεων σε διαφορετικά επίπεδα αφαίρεσης. Κατά συνέπεια, σε εργασίες αναγνώρισης που αποτελούνται από εκατοντάδες κατηγορίες εικόνων, οι βαθιές CNN έχουν παρουσιάσει σημαντική βελτίωση των επιδόσεων σε σχέση με τα συμβατικά μοντέλα που βασίζονται στην όραση (Ojala et al. 2002- Dalal and Triggs 2004- Lowe 2004). Η παρατήρηση ότι τα βαθιά αρχιτεκτονικές μπορούν να βελτιώσουν την αναπαραστατική ικανότητα ενός CNN ενίσχυσε τη χρήση των CNN σε ταξινόμηση εικόνων και εργασίες κατάτμησης (Krizhevsky et al. 2012). Η διαθεσιμότητα μεγάλων δεδομένων και οι εξελίξεις στο υλικό είναι επίσης οι κύριοι λόγοι για την πρόσφατη επιτυχία των βαθιών CNN. Εμπειρικές μελέτες έδειξαν ότι αν δοθούν αρκετά δεδομένα εκπαίδευσης, τα βαθιά CNNs μπορούν να μάθουν το αναλλοίωτο αναπαραστάσεις και μπορούν να επιτύχουν επιδόσεις ανθρώπινου επιπέδου. Εκτός από τη χρήση τους ως εποπτευόμενα μηχανισμό μάθησης, οι δυνατότητες των βαθιών CNNs μπορούν επίσης να αξιοποιηθούν για την εξαγωγή χρήσιμων αναπαραστάσεις από μεγάλη κλίμακα δεδομένων χωρίς ετικέτες. Πρόσφατα, έχει αποδειχθεί ότι διαφορετικά επίπεδα χαρακτηριστικών, συμπεριλαμβανομένων τόσο χαμηλού όσο και υψηλού επιπέδου, μπορούν να μεταφερθούν σε μια γενική εργασία αναγνώρισης μέσω αξιοποιώντας την έννοια της μάθησης μεταφοράς

Από τα τέλη της δεκαετίας του 1990 έως το 2000, διάφορες βελτιώσεις στη μεθοδολογία εκμάθησης CNN και της αρχιτεκτονικής πραγματοποιήθηκαν για να καταστήσουν το CNN κλιμακούμενο σε μεγάλες, ετερογενείς, σύνθετες και προβλήματα πολλαπλών κατηγοριών. Οι καινοτομίες στα CNN περιλαμβάνουν διάφορες πτυχές, όπως η τροποποίηση των μονάδων επεξεργασίας, στρατηγικές βελτιστοποίησης παραμέτρων και υπερ-παραμέτρων, πρότυπα σχεδιασμού και

συνδεσιμότητα των στρωμάτων, κ.λπ. Οι εφαρμογές που βασίζονται σε CNN επικράτησαν μετά την υποδειγματική απόδοση του AlexNet στο σύνολο δεδομένων ImageNet το 2012 (Krizhevsky et al. 2012). Σημαντικές καινοτομίες στα CNN έχουν

προταθεί από τότε και αποδίδονται σε μεγάλο βαθμό στην αναδιάρθρωση των μονάδων επεξεργασίας και στο σχεδιασμό νέων μπλοκ. Zeiler και Fergus (Zeiler and Fergus 2013) έδωσαν την έννοια της απεικόνισης του CNN κατά στρώματα για τη βελτίωση της κατανόησης των χαρακτηριστικών σταδίων εξαγωγής, η οποία μετατόπισε την τάση προς την εξαγωγή χαρακτηριστικών σε χαμηλή χωρική ανάλυση στη βαθιά αρχιτεκτονική, όπως πραγματοποιείται στο VGG (Simonyan and Zisserman 2015). Στις μέρες μας, τα περισσότερα από τα νέες αρχιτεκτονικές βασίζονται στην αρχή της απλής και ομοιογενούς τοπολογίας, όπως εισήχθη στην VGG. Η ομάδα βαθιάς μάθησης της Google εισήγαγε μια καινοτόμο ιδέα της διάσπασης, μετασχηματισμού και συγχώνευσης, με το αντίστοιχο μπλοκ να είναι γνωστό ως μπλοκ έναρξης. Το inception μπλοκ έδωσε για πρώτη φορά την έννοια της διακλάδωσης εντός ενός επιπέδου, η οποία επιτρέπει αφαίρεση χαρακτηριστικών σε διαφορετικές χωρικές κλίμακες. Το 2015, η έννοια της

συνδέσεων παράλειψης που εισήγαγε το ResNet (He et al. 2015a) για την εκπαίδευση βαθιών CNN απέκτησε δημοτικότητα. Στη συνέχεια, η έννοια αυτή χρησιμοποιήθηκε από τα περισσότερα από τα επόμενα δίκτυα, όπως τα Inception-ResNet, Wide ResNet, ResNeXt κ.λπ.

Διαφορετικά αρχιτεκτονικά σχέδια, όπως Wide ResNet, ResNeXt, Pyramidal Net, Xception, PolyNet, και πολλά άλλα διερευνούν την επίδραση των πολυεπίπεδων μετασχηματισμών στη μάθηση των CNNs ικανότητα εισάγοντας την καρδινικότητα ή αυξάνοντας το πλάτος (Zagoruyko and Komodakis 2016,

Han et al. 2017- Xie et al. 2017- Zhang et al. 2017). Ως εκ τούτου, το επίκεντρο της έρευνας μετατοπίστηκε από τη βελτιστοποίηση παραμέτρων και την αναπροσαρμογή συνδέσεων προς τη βελτίωση του αρχιτεκτονικού σχεδιασμού του δικτύου. Αυτή η μετατόπιση οδήγησε σε πολλές νέες αρχιτεκτονικές ιδέες, όπως η ενίσχυση καναλιών, χωρική και αξιοποίηση με βάση το χάρτη χαρακτηριστικών και επεξεργασία πληροφοριών με βάση την προσοχή κ.λπ. Τα τελευταία χρόνια διεξάγονται διάφορες ενδιαφέρουσες έρευνες σχετικά με τα βαθιά CNN που επεξεργάζονται τα βασικά στοιχεία των CNN και τις εναλλακτικές τους λύσεις.

YOLO(YOU ONLY LOOK ONCE)

Περίληψη:

Η ανίχνευση αντικειμένων παίζει θεμελιώδη ρόλο στις εφαρμογές υπολογιστικής όρασης, από την αυτόνομη οδήγηση μέχρι τα συστήματα επιτήρησης. Μεταξύ διαφόρων αλγορίθμων ανίχνευσης αντικειμένων, το πλαίσιο YOLO (You Only Look Once) έχει συγκεντρώσει σημαντική προσοχή για την αξιοσημείωτη ισορροπία μεταξύ ταχύτητας και ακρίβειας. Στην παρούσα εργασία, παρέχουμε μια εμπεριστατωμένη ανασκόπηση του πλαισίου YOLO, καλύπτοντας την εξέλιξή του, την αρχιτεκτονική του, τις στρατηγικές εκπαίδευσης και τις εφαρμογές του στον πραγματικό κόσμο. Συζητάμε επίσης τις πρόσφατες εξελίξεις και παραλλαγές του YOLO, μαζί με τις

προκλήσεις και τις μελλοντικές κατευθύνσεις στον τομέα της ανίχνευσης αντικειμένων.

1. Εισαγωγή:

Η ανίχνευση αντικειμένων είναι μια κρίσιμη εργασία στην όραση υπολογιστών, με στόχο τον εντοπισμό και την ταξινόμηση αντικειμένων μέσα σε εικόνες ή βίντεο. Οι παραδοσιακές προσεγγίσεις ανίχνευσης αντικειμένων περιλαμβάνουν συνήθως σωληνώσεις πολλαπλών σταδίων, οι οποίες μπορεί να είναι υπολογιστικά δαπανηρές και αργές, περιορίζοντας τη δυνατότητα εφαρμογής τους σε σενάρια πραγματικού χρόνου. Το YOLO, που παρουσιάστηκε από τους Joseph Redmon et al. το 2016, έφερε επανάσταση στην ανίχνευση αντικειμένων προτείνοντας μια ενοποιημένη προσέγγιση που επεξεργάζεται ολόκληρη την εικόνα ταυτόχρονα, παρέχοντας απόδοση σε πραγματικό χρόνο χωρίς συμβιβασμούς στην ακρίβεια. Από την ίδρυσή του, το πλαίσιο YOLO έχει υποστεί αρκετές επαναλήψεις και έχει εμπνεύσει πολυάριθμες εξελίξεις, καθιστώντας το έναν από τους αλγορίθμους με τη μεγαλύτερη επιρροή στον τομέα της όρασης υπολογιστών.

2. Εξέλιξη του YOLO:

Το πλαίσιο YOLO έχει εξελιχθεί σημαντικά από την αρχική του δημιουργία. Το YOLOv1 εισήγαγε την έννοια της διαίρεσης της εικόνας εισόδου σε ένα πλέγμα και της ταυτόχρονης πρόβλεψης οριοθετημένων πλαισίων και πιθανοτήτων κλάσης για κάθε κελί πλέγματος. Οι μεταγενέστερες εκδόσεις, όπως οι YOLOv2 και YOLOv3, ενσωμάτωσαν βελτιώσεις στην αρχιτεκτονική και τις στρατηγικές εκπαίδευσης, με αποτέλεσμα καλύτερη ακρίβεια και αποδοτικότητα. Το YOLOv4 εισήγαγε νέες τεχνικές για την περαιτέρω βελτιστοποίηση της ταχύτητας και της ακρίβειας, ενώ το YOLOv5 επικεντρώθηκε σε ελαφρές αρχιτεκτονικές και εξορθολογισμένες σωληνώσεις εκπαίδευσης.

3. Αρχιτεκτονική του YOLO:

Η αρχιτεκτονική του YOLO αποτελείται από μια ραχοκοκαλιά νευρωνικού δικτύου συνελκτικού τύπου (CNN), ακολουθούμενη από στρώματα ανίχνευσης που είναι υπεύθυνα για την πρόβλεψη των οριοθετημένων πλαισίων και των πιθανοτήτων κλάσης. Η ραχοκοκαλιά CNN εξάγει χαρακτηριστικά από την εικόνα εισόδου, τα οποία στη συνέχεια περνούν από πολλαπλά στρώματα ανίχνευσης για τη δημιουργία προβλέψεων. Το YOLO χρησιμοποιεί ένα ενιαίο νευρωνικό δίκτυο για την εκτέλεση τόσο του εντοπισμού όσο και της ταξινόμησης, επιτρέποντας την εκπαίδευση και την εξαγωγή συμπερασμάτων από άκρη σε άκρη.

4. Στρατηγικές εκπαίδευσης:

Η εκπαίδευση ενός μοντέλου YOLO περιλαμβάνει τη βελτιστοποίηση διαφόρων συνιστωσών, συμπεριλαμβανομένων των συναρτήσεων απώλειας, των τεχνικών επαύξησης δεδομένων και των υπερπαραμέτρων. Το YOLO χρησιμοποιεί έναν συνδυασμό απωλειών εντοπισμού και απωλειών ταξινόμησης για την εκπαίδευση του μοντέλου, με πρόσθετες τεχνικές κανονικοποίησης, όπως η ομαλοποίηση παρτίδας και η εγκατάλειψη. Η επαύξηση δεδομένων παίζει καθοριστικό ρόλο στη βελτίωση της ευρωστίας του μοντέλου με την εισαγωγή διαφοροποιήσεων στα δεδομένα εκπαίδευσης, ενώ η ρύθμιση των υπερπαραμέτρων συμβάλλει στη βελτιστοποίηση της απόδοσης του μοντέλου σε συγκεκριμένες εργασίες και σύνολα δεδομένων.

5. Πραγματικές εφαρμογές:

Το πλαίσιο YOLO έχει βρει ευρέως διαδεδομένες εφαρμογές σε διάφορους τομείς, όπως η ανίχνευση αντικειμένων σε εικόνες και βίντεο, η ανίχνευση πεζών για συστήματα αυτόνομης οδήγησης, η ανίχνευση προσώπων σε υλικό παρακολούθησης και η παρακολούθηση αντικειμένων σε περιβάλλοντα πραγματικού χρόνου. Η ικανότητά του να επιτυγχάνει επιδόσεις σε πραγματικό χρόνο, διατηρώντας παράλληλα ανταγωνιστική ακρίβεια, το καθιστά ελκυστική επιλογή για την ανάπτυξη συστημάτων ανίχνευσης αντικειμένων σε περιβάλλοντα με περιορισμένους πόρους.

6. Εξελίξεις και παραλλαγές:

Έχουν προταθεί διάφορες εξελίξεις και παραλλαγές του πλαισίου YOLO για την αντιμετώπιση συγκεκριμένων προκλήσεων ή τη βελτιστοποίηση των επιδόσεων για διαφορετικές περιπτώσεις χρήσης. Αυτές περιλαμβάνουν τα YOLO Nano, YOLO-LITE και YOLObile, τα οποία επικεντρώνονται σε συμπαγείς αρχιτεκτονικές για συσκευές άκρων και κινητές πλατφόρμες, καθώς και εξειδικευμένες παραλλαγές προσαρμοσμένες για συγκεκριμένους τομείς, όπως η ανάλυση αεροφωτογραφιών και η ανίχνευση οχημάτων.

7. Προκλήσεις και μελλοντικές κατευθύνσεις:

Παρά την επιτυχία του, το πλαίσιο YOLO εξακολουθεί να αντιμετωπίζει προκλήσεις, όπως ο χειρισμός μικρών ή πυκνά στοιβαγμένων αντικειμένων, η αντιμετώπιση αποκρύψεων και ακατάστατων σκηνών και η επίτευξη ανθεκτικότητας στις μεταβολές του φωτισμού και των περιβαλλοντικών συνθηκών. Οι μελλοντικές ερευνητικές κατευθύνσεις περιλαμβάνουν τη διερεύνηση νέων αρχιτεκτονικών σχεδίων, την ενσωμάτωση πληροφοριών σχετικά με το πλαίσιο και την αξιοποίηση των εξελίξεων στη μάθηση χωρίς επίβλεψη και την αυτοεπιτήρηση για την περαιτέρω βελτίωση των επιδόσεων των συστημάτων ανίχνευσης αντικειμένων με βάση το YOLO.

8. Συμπεράσματα:

Συμπερασματικά, το πλαίσιο YOLO έχει προωθήσει σημαντικά τον τομέα της ανίχνευσης αντικειμένων παρέχοντας μια ενιαία και αποτελεσματική λύση για απόδοση σε πραγματικό χρόνο. Η εξέλιξή του, από το αρχικό YOLOv1 έως τις τελευταίες παραλλαγές, καταδεικνύει συνεχή καινοτομία και βελτίωση τόσο της ταχύτητας όσο και της ακρίβειας. Με την ευρεία υιοθέτησή του και τις συνεχιζόμενες ερευνητικές προσπάθειες, το YOLO είναι έτοιμο να παραμείνει στην πρώτη γραμμή της τεχνολογίας ανίχνευσης αντικειμένων, οδηγώντας σε εξελίξεις στις εφαρμογές υπολογιστικής όρασης σε διάφορους τομείς.

Image Segmentation

Περίληψη

Ο διαχωρισμός θεωρείται ως ένα από τα κύρια βήματα στην επεξεργασία εικόνας. Διαιρεί μια ψηφιακή εικόνα σε σε πολλαπλές περιοχές προκειμένου να τις αναλύσει. Χρησιμοποιείται επίσης για τη διάκριση διαφορετικών αντικειμένων στην εικόνα. Αρκετές εικόνες τεχνικές τμηματοποίησης έχουν αναπτυχθεί από τους ερευνητές προκειμένου να καταστήσουν τις εικόνες ομαλές και εύκολα να αξιολογούνται.

ΕΙΣΑΓΩΓΗ

Διάσημες τεχνικές τμηματοποίησης εικόνας οι οποίες εξακολουθούν να χρησιμοποιούνται από τους ερευνητές είναι η ανίχνευση ακμών, το κατώφλι, τα ιστογράμματα, μέθοδοι που βασίζονται στην περιοχή και Watershed Μετασχηματισμός. Δεδομένου ότι οι εικόνες χωρίζονται σε δύο τύπους με βάση το χρώμα τους, δηλαδή σε εικόνες κλίμακας του γκρι και έγχρωμες εικόνες. Ως εκ τούτου, η κατάτμηση εικόνων για έγχρωμες εικόνες είναι εντελώς διαφορετική από τις εικόνες κλίμακας του γκρι, π.χ. η τμηματοποίηση εικόνων με βάση το περιεχόμενο. ανάκτηση εικόνας. Επίσης, ποιος αλγόριθμος είναι εύρωστος και λειτουργεί καλά εξαρτάται από τον τύπο της εικόνας. Η ιδιότητα ενός εικονοστοιχείου σε μια εικόνα και οι πληροφορίες των εικονοστοιχείων κοντά σε αυτό το εικονοστοιχείο είναι δύο βασικές παράμετροι για κάθε τμηματοποίηση εικόνας αλγόριθμο. Μπορεί επίσης να αναπαρασταθεί ως ομοιότητα των εικονοστοιχείων σε οποιαδήποτε περιοχή και ασυνέχεια των ακμών στην εικόνα. Με βάση τις ακμές χρησιμοποιείται για τη διαίρεση της εικόνας με βάση τις ακμές τους ακμών. Οι μέθοδοι με βάση την περιοχή χρησιμοποιούν το κατώφλι προκειμένου να διαχωρίσουν το φόντο από μια εικόνα, ενώ οι νευρωνικές τεχνικές που βασίζονται σε δίκτυα χρησιμοποιούν τον αλγόριθμο μάθησης για την εκπαίδευση τη διαδικασία τμηματοποίησης της εικόνας [4]. Το αποτέλεσμα που λαμβάνεται από διαδικασία τμηματοποίησης εικόνας είναι η κύρια παράμετρος για την περαιτέρω αυτό το αποτέλεσμα θα καθορίσει επίσης την την

ποιότητα της περαιτέρω διαδικασίας επεξεργασίας εικόνας. Οι αλγόριθμοι τμηματοποίησης παίζουν σημαντικό ρόλο στις ιατρικές εφαρμογές, δηλαδή στη διάγνωση ασθενειών που σχετίζονται με τον εγκέφαλο την καρδιά, το γόνατο, τη σπονδυλική στήλη, τη λεκάνη, τον προστάτη και τα αιμοφόρα αγγεία, και εντοπισμό παθολογίας. Ως εκ τούτου, η τμηματοποίηση εικόνας εξακολουθεί να ένας πολύ καυτός τομέας έρευνας για τον τομέα της επεξεργασίας εικόνας. Είναι ένα δύσκολο έργο για τους ερευνητές και τους προγραμματιστές να αναπτύξουν μια καθολική τεχνική για την τμηματοποίηση εικόνας. Η τμηματοποίηση εικόνας χρησιμοποιείται επίσης για τη διάκριση διαφορετικών αντικειμένων στην εικόνα, αφού η εικόνα μας χωρίζεται στο προσκήνιο και το φόντο, ενώ το προσκήνιο της εικόνας είναι σχετιζόμενο με την περιοχή ενδιαφέροντος, καθώς το φόντο είναι το υπόλοιπο τμήμα της εικόνας.. Ως εκ τούτου, η τμηματοποίηση εικόνας θα διαχωρίσει αυτά τα δύο μέρη μεταξύ τους. Οι τρεις κύριες προσεγγίσεις τμηματοποίησης της εικόνας είναι το Κατώφλι, η Ακμή και η σχετιζόμενη με την περιοχή.

A. Εξαγωγή Εικόνας Βάσει Ακμών

Ο Fernando C. Monteiro [πρότεινε μια νέα μέθοδο εξαγωγής εικόνας που αποτελείται από πληροφορίες βάσει ακμών και περιοχής με τη βοήθεια της φασματικής μεθόδου και μορφολογικού αλγορίθμου watershed. Καταρχάς, μειώνουν το θόρυβο από την εικόνα χρησιμοποιώντας φιλτράρισμα διπλής πλευράς ως ένα βήμα προεπεξεργασίας. Δεύτερον, χρησιμοποιείται συγχώνευση περιοχής για να πραγματοποιηθεί προκαταρκτική εξαγωγή. Επιπλέον δημιουργείται ομοιότητα περιοχής και στη συνέχεια πραγματοποιείται ομαδοποίηση περιοχής βάσει γράφου χρησιμοποιώντας τη μέθοδο Multi-class Normalized Cut . Χρησιμοποιείται το σύνολο δεδομένων Berkeley segmentation dataset. Συγκρίνουν την τεχνική με τη μέθοδο mean shift, πολυκλιμακική ομαδοποίηση βάσει γράφου και JSEG. Παρατηρήθηκε ότι η προτεινόμενη τεχνική ξεπέρασε τις άλλες μεθόδους και παρήγαγε καλύτερα αποτελέσματα.

Ο R. V. Patil ισχυρίζεται ότι εάν ο αριθμός των συστάδων εκτιμηθεί με ακρίβεια, η κατάτμηση εικόνας με τη μέθοδο K-means θα παρέχει καλύτερα αποτελέσματα. Πρότεινε μια νέα μέθοδο που βασίζεται στην ανίχνευση ακμών για την εκτίμηση του αριθμού των συστάδων. Η σύμπτωση φάσης χρησιμοποιείται για την ανίχνευση των ακμών. Στη συνέχεια, αυτές οι ακμές χρησιμοποιούνται για την εύρεση συστάδων. Χρησιμοποιείται κατώφλι και ευκλείδεια απόσταση για τη δημιουργία συστάδων. Για την εύρεση της τελικής τμηματοποίησης της εικόνας χρησιμοποιείται το K-means. Για την υλοποίηση της προτεινόμενης τεχνικής χρησιμοποιείται το MATLAB. Πραγματοποιούνται πειράματα σε εννέα διαφορετικές εικόνες και τα αποτελέσματα δείχνουν ότι ο αριθμός των συστάδων είναι ακριβής και βέλτιστος. Ο Weihong Cui Yi Zhang πρότεινε μια μέθοδο αυτόματης επιλογής κατωφλίου με βάση τις ακμές για τη δημιουργία τμηματοποίησης εικόνας πολλαπλών κλιμάκων. Για τον υπολογισμό του βάρους των ακμών χρησιμοποιείται το βάρος της ζώνης και ο NDVI (Normalized Difference Vegetation Index). Η μέθοδος MST και η μέθοδος κατωφλίου με βάση την ακμή χρησιμοποιείται για την τμηματοποίηση της εικόνας.

Τα πειράματα εκτελούνται σε εικόνες ανάλυσης πολλαπλών κλιμάκων, δηλαδή σε πολυφασματικές εικόνες Quick-bird. Τα αποτελέσματα έδειξαν ότι η μέθοδός τους διατηρεί τις πληροφορίες για το αντικείμενο και τα όρια του αντικειμένου κατά την κατάτμηση της εικόνας. Η Anna Fabijańska εισήγαγε μια νέα μέθοδο που χρησιμοποιεί φίλτρο διακύμανσης για την ανίχνευση ακμών στη διαδικασία τμηματοποίησης εικόνας. Η μέθοδός τους βρήκε τη θέση της ακμής χρησιμοποιώντας φίλτρο διακύμανσης. Το φίλτρο Sobel Gradient με K-means χρησιμοποιείται επίσης για την εξαγωγή των ακμών και συγκρίνεται με την προτεινόμενη τεχνική. Συζητείται επίσης η επίδραση του μεγέθους του παραθύρου φιλτραρίσματος στον προσδιορισμό των ακμών και διαπιστώνεται ότι εάν το παράθυρο 9×9 χρησιμοποιείται για την εξαγωγή των ακμών τότε η ακμή είναι πλήρης και ταιριάζει με ακρίβεια με το σχήμα του αντικειμένου στην εικόνα. Στην περίπτωση εικόνων με μεγαλύτερες λεπτομέρειες, προτείνεται ένα μικρό παράθυρο φιλτραρίσματος. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη τεχνική ξεπερνά τον ανιχνευτή ακμών Sobel.

Ο Mohammed J. Islam διαπίστωσε ότι η υπολογιστική όραση είναι η καλύτερη λύση για την επιθεώρηση καψουλών σε πραγματικό χρόνο στη φαρμακευτική βιομηχανία. Ο συγγραφέας ανέπτυξε ένα σύστημα για την επιθεώρηση της ποιότητας χρησιμοποιώντας τεχνικές κατάτμησης εικόνας με βάση τις ακμές. Χρησιμοποίησαν τον ανιχνευτή ακμών Sobel [18] προκειμένου να ανιχνεύσουν τις ακμές με την ιδιότητα καταστολής του θορύβου. Μετά την ανίχνευση ακμών, χρησιμοποιείται η τεχνική Otsu Thresholding για τον εντοπισμό των εικονοστοιχείων φόντου και προσκηνίου. Διεξάγονται πειράματα και τα αποτελέσματα συγκρίνονται με την τεχνική τμηματοποίησης με βάση το NN που κατασκευάζει το Visual C++. Τα αποτελέσματα υπερτερούν της τεχνικής NN με βάση την ακρίβεια και τη διαφορά χρόνου επεξεργασίας 10 ms.

4. Κατάτμηση εικόνας με βάση την ασαφή θεωρία

Ο Liu Yucheng πρότεινε έναν νέο αλγόριθμο τμηματοποίησης εικόνων με βάση την ασαφή μορφολογία συγχώνευσης. Ο αλγόριθμος χρησιμοποίησε μορφολογικές λειτουργίες ανοίγματος και κλεισίματος για την εξομάλυνση της εικόνας και στη συνέχεια εκτέλεσε τις λειτουργίες κλίσης στην προκύπτουσα εικόνα. Μετά τη σύγκριση του προτεινόμενου αλγορίθμου σύντηξης με τον αλγόριθμο Watershed και τις μεθόδους Prewitt, διαπιστώνεται ότι η προσέγγιση σύντηξης επιλύει το πρόβλημα της υπερτμηματοποίησης του αλγορίθμου Watershed. Σώζει επίσης τις λεπτομέρειες των πληροφοριών της εικόνας και βελτιώνει επίσης την ταχύτητα. Ο Syoji Kobashi χρησιμοποίησε την κατάτμηση εικόνας με βάση την κλίμακα και το ασαφές μοντέλο αντικειμένου για την κατάτμηση της περιοχής του εγκεφαλικού παρεγχύματος σε εικόνα μαγνητικής τομογραφίας εγκεφάλου νεογέννητου. Η περιοχή του προσκηνίου διαχωρίζεται στο πρώτο βήμα, στη συνέχεια πραγματοποιείται διόρθωση της ανομοιογένειας της έντασης της μαγνητικής τομογραφίας και στη συνέχεια εφαρμόζεται το ασαφές μοντέλο αντικειμένου με βάση την κλίμακα (FOM) στην προκύπτουσα εικόνα. Τα αποτελέσματα της προτεινόμενης μεθόδου αξιολογούνται με βάση το γρήγορο θετικό κλάσμα όγκου (FPVF) και το γρήγορο αρνητικό κλάσμα όγκου (FVNF). Τα αποτελέσματα του πειράματος έδειξαν ότι το FOM (ασαφές μοντέλο αντικειμένου) πέτυχε τις ελάχιστες τιμές FPVF και FVNF. Ο Refik Samet πρότεινε μια νέα τεχνική τμηματοποίησης εικόνας με βάση ασαφείς κανόνες για την τμηματοποίηση των εικόνων λεπτών τμημάτων βράχων. Παίρνουν την εικόνα RGB

του λεπτού τμήματος πετρώματος ως είσοδο και δίνουν την τμηματοποιημένη εικόνα ορυκτών ως έξοδο. Το Fuzzy C Means εφαρμόζεται επίσης σε εικόνες λεπτού πετρώματος και τα αποτελέσματα συγκρίνονται και των δύο τεχνικών. Αρχικά, ο χρήστης θα λάβει δείγμα εικόνας από ορυκτά- τα χαρακτηριστικά διακρίνονται με βάση τις κόκκινες, πράσινες και μπλε συνιστώσες της εικόνας. Η συνάρτηση συμμετοχής ορίζεται για κάθε συνιστώσα χρησιμοποιώντας ασαφείς κανόνες. Κάθε συνάρτηση συμμετοχής αντιπροσωπεύει την κατανομή του χρώματος στην εικόνα. Ορίζονται ισχυρά και αδύναμα σημεία, ενώ τα ισχυρά σημεία θεωρούνται σημεία εκκίνησης και τα αδύναμα σημεία γίνονται μέλη τους. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη τεχνική είναι καλύτερη από τον αλγόριθμο FCM.

. Ο Muhammad Rizwan Khokher

παρουσίασε μια νέα μέθοδο τμηματοποίησης εικόνας με χρήση συστήματος βασισμένου σε ασαφείς κανόνες και Graph Cuts. Οι επιστήμονες τμηματοποίησαν αρχικά τις εικόνες κλίμακας του γκρι χρώματος και υφής χρησιμοποιώντας Graph Cuts. Τα βάρη αποδίδονται στα χαρακτηριστικά της εικόνας χρησιμοποιώντας ασαφείς κανόνες. Ο αλγόριθμός τους λειτουργεί αρχικά με την εξαγωγή των χαρακτηριστικών της εικόνας, τον υπολογισμό των σταθερών με τη χρήση ασαφών κανόνων, τον υπολογισμό του σταθμισμένου μέσου όρου των σταθερών για την εύρεση του πίνακα ομοιότητας, την κατάτμηση του γραφήματος με τη μέθοδο Normalized Graph Cut και τέλος την τμηματοποιημένη εικόνα από το κατατμημένο γράφημα. Η βάση δεδομένων Berkley χρησιμοποιείται για την αξιολόγηση του αλγορίθμου. Η προσομοίωση πραγματοποιείται σε Matlab και γλώσσα C. Τα αποτελέσματα αξιολογούνται με βάση τη μέση τιμή, την τυπική απόκλιση και την τιμή PPV. Διαπιστώνεται ότι τα περισσότερα από τα μέτρα αξιολόγησης κυμαίνονται από 0,85 έως 0,95 για S.D και PPV.

Α. Τμηματοποίηση εικόνας με βάση την Μερική Διαφορική Εξίσωση (PDE)

Ο Jinsheng Xiao πρότεινε μια νέα μη γραμμική διακοπτόμενη μερική διαφορική εξίσωση (PDE) που μοντελοποιεί τη μέθοδο level set των γκριζών εικόνων. Προτείνεται επίσης μια διακριτή μέθοδος για την εύρεση αριθμητικής λύσης και την υλοποίηση του φίλτρου. Ο τύπος της μη γραμμικής ασυνεχούς PDE εφαρμόζεται σε εικόνα εικονολήπτη με χρήση του MATLAB. Τα αποτελέσματα έδειξαν ότι οι ακμές και τα όρια της εικόνας παραμένουν θολά και μπορούν να μετατοπιστούν με τη χρήση του τελεστή Close. Περισσότερες πληροφορίες μπορούν να αποθηκευτούν με τη χρήση του προτεινόμενου σχήματος. Ο Fengchun Zhang [2] παρουσιάζει ένα μοντέλο παραλλαγής που χρησιμοποιεί PDE 4ης τάξης με δεύτερη PDE τάξης για την αποθορυβοποίηση εικόνων φλεβών δακτύλων. Η τεχνική τμηματοποίησης Midpoint Threshold χρησιμοποιείται για την ακριβή εξαγωγή της περιοχής ενδιαφέροντος. Η PDE τέταρτης τάξης μείωσε πολύ καλά το θόρυβο, ενώ η PDE δεύτερης τάξης προσέγγισε αποτελεσματικά τα όρια. Από τα πειράματα παρατηρείται ότι η τιμή PSNR της προτεινόμενης μεθόδου αυξάνεται κατά 2 dB. Η μέθοδος συγκρίνεται με τον αλγόριθμο κατάτμησης με βάση το κατώφλι και διαπιστώνεται ότι η προτεινόμενη μέθοδος κατάτμησης με ακρίβεια την πραγματική εικόνα φλεβών δακτύλου. Ο Chun Yuan πρότεινε ένα νέο μοντέλο τμηματοποίησης για έγχρωμες εικόνες. Το μοντέλο βασίζεται στο μοντέλο Geodesic Active Contour (GAC). Αλλά

το GAC περιορίζεται μόνο σε εικόνες κλίμακας του γκρι. Ως εκ τούτου, το μοντέλο τους αποτελεί επίσης επέκταση του μοντέλου GAC και είναι γνωστό ως μοντέλο έγχρωμου GAC. Χρησιμοποιεί την έκφραση της κλίσης της έγχρωμης εικόνας.

Τεχνητό νευρωνικό δίκτυο (ANN) με βάση την κατάτμηση εικόνας

Ο Wencang Zhao πρότεινε έναν νέο αλγόριθμο τμηματοποίησης εικόνων που βασίζεται σε χαρακτηριστικά υφής και Νευρωνικό Δίκτυο για τον διαχωρισμό των στοχευμένων εικόνων από το φόντο. Χρησιμοποιείται σύνολο δεδομένων από εικόνες μικρο-CT. Χρησιμοποιείται φίλτρο αποθρομβοποίησης για την αφαίρεση του θορύβου από την εικόνα ως βήμα προεπεξεργασίας, στη συνέχεια πραγματοποιείται εξαγωγή χαρακτηριστικών και στη συνέχεια δημιουργείται νευρωνικό δίκτυο διάδοσης πίσω (Back Propagation Neural Network) και, τέλος, τροποποιεί τον αριθμό βαρών του δικτύου και αποθηκεύει την έξοδο. Ο προτεινόμενος αλγόριθμος συγκρίνεται με τη μέθοδο Thresholding και τη μέθοδο Region Growing. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη τεχνική υπερτερεί των άλλων μεθόδων με βάση την ταχύτητα και την ακρίβεια της κατάτμησης. Ο Lijun Zhang πρότεινε ένα νέο σύστημα τμηματοποίησης εικόνων με βάση νευρωνικά δίκτυα για έγχρωμες εικόνες. Συνδύασε την αποσύνθεση Wavelet και τον αυτοοργανωτικό χάρτη (SOM) για να προτείνει μια νέα μέθοδο, δηλαδή το SOM-NN. Η ψηφοφορία μεταξύ των εικονοστοιχείων-παιδιών επέλεξε το εικονοστοιχείο-γονέα. Μετά την αρχικοποίηση, η ANN βρήκε το αποτέλεσμα τμηματοποίησης που ικανοποιεί όλα τα επίπεδα. Πραγματοποιείται αποσύνθεση Wavelet για την απομάκρυνση του θορύβου. Ως εκ τούτου, η αποσύνθεση wavelet και το SOM-NN συνδυάζονται για την εκτέλεση της τμηματοποίησης. Τα αποτελέσματα έδειξαν ότι η μέθοδος έχει μειώσει το θόρυβο και παράγει ακριβή τμηματοποίηση. Ο Shohel Ali Ahmed πρότεινε τεχνική ταξινόμησης υφής εικόνας με βάση τεχνητά νευρωνικά δίκτυα (ANN). Αρχικά, η εικόνα συλλαμβάνεται και πραγματοποιείται προεπεξεργασία, μετά από αυτό, πραγματοποιείται εξαγωγή χαρακτηριστικών, ενώ ο ταξινομητής ANN χρησιμοποιείται για την ταξινόμηση υφής, πραγματοποιείται ομαδοποίηση για το διαχωρισμό του φόντου από τις υποεικόνες. Το εκπαιδευμένο ANN συνδυάζει τα εικονοστοιχεία εισόδου σε δύο συστάδες που δίνουν αποτελέσματα. Παράγει την ταξινόμηση υφής και την κατάτμηση της εικόνας.

Τμηματοποίηση εικόνας με βάση το κατώφλι

Ο Shiping Zhu πρότεινε έναν νέο αλγόριθμο ανίχνευσης ακμών και τμηματοποίησης εικόνας με βάση το κατώφλι. Υπολογίζει το κατώφλι κάθε εικονοστοιχείου στην εικόνα με βάση τα γειτονικά του εικονοστοιχεία. Βρίσκει επίσης τις ακμές της εικόνας με τη βοήθεια του προτεινόμενου αλγορίθμου. Το κατώφλι κάθε εικονοστοιχείου ορίστηκε με τη χρήση ιστογράμματος. Το PDF χρησιμοποιείται για την απομόνωση του φόντου και του κατωφλίου της εικόνας. Υλοποιούν τον αλγόριθμό τους σε Visual C++. Τα αποτελέσματα ξεπερνούν τα αποτελέσματα του Canny Operator, επειδή εκτελεί ταυτόχρονα ανίχνευση ακμών και τμηματοποίηση. Ο Anping XU πρότεινε μια προσέγγιση που βασίζεται σε σύνολο επιπέδων με βάση το κατώφλι και περιλαμβάνει

τόσο την τμηματοποίηση με βάση το κατώφλι όσο και τη μέθοδο γρήγορης πορείας (FMM) για την τμηματοποίηση ιατρικών εικόνων. Το αποτέλεσμα του φίλτρου αποθορυβοποίησης περνά στην FMM για σκοπούς τμηματοποίησης με τη βοήθεια της τεχνικής level set με βάση το κατώφλι. Υλοποιούν τη μεθοδό τους σε VC++ και ITK. Μετά το πείραμα, τα αποτελέσματα έδειξαν ότι η μέθοδος level set με βάση το κατώφλι οδηγεί σε σαφέστερη, ακριβέστερη και τελειότερη τμηματοποίηση, αποφεύγει επίσης τη διαρροή ακμών και βελτιώνει την ταχύτητα της τμηματοποίησης.

Οι Wu Kaihua και Ban Tao [παρουσίασαν μια νέα μέθοδο τμηματοποίησης βέλτιστου κατωφλίου βασισμένη σε κριτήρια εντροπίας και γενετικό αλγόριθμο, προκειμένου να βελτιώσουν τη διαδικασία λήψης εικόνας στην όραση υπολογιστών. Οι παράγοντες που λαμβάνονται από αυτούς είναι ο φωτισμός, το φως, η αντανάκλαση του φωτός, ο χρόνος έκθεσης του CCD και ορισμένες ιδιότητες του ιστογράμματος της εικόνας. Συγκρίνουν την προτεινόμενη τεχνική τους με τον αλγόριθμο Otsu και διαπίστωσαν ότι ο αλγόριθμός τους είναι αποτελεσματικός στην αναζήτηση και στην εύρεση τμηματοποίησης μιας εικόνας βάσει κατωφλίου. Ο Frank Jiang πρότεινε μια νέα πολυεπίπεδη τεχνική τμηματοποίησης με βάση το κατώφλι χρησιμοποιώντας PSO και μετάλλαξη Wavelet. Πρότειναν επίσης έναν νέο αλγόριθμο PSO ο οποίος χρησιμοποιείται στα δύο πρώτα βήματα του αλγορίθμου. Στη συνέχεια, η έξοδος του PSO περνάει στις λειτουργίες μετάλλαξης wavelet, οι οποίες εκτελούν τη λειτουργία μετάλλαξης και ενημερώνουν τις εξισώσεις PSO μετά από αυτή. Αυτή η εργασία θα δημιουργήσει βελτιστοποιημένο κατώφλι και σωστή τμηματοποίηση. Μετά από σύγκριση της μεθόδου τους με την HCOCLPSO, διαπίστωσαν ότι παράγει βέλτιστο κατώφλι σε σύγκριση με την άλλη μέθοδο. Ισχυρίζονται ότι ο αλγόριθμός τους είναι ο καλύτερος για εφαρμογές πραγματικού χρόνου, π.χ. εφαρμογή βίντεο ανθεκτική σε σφάλματα σε εχθρικό περιβάλλον.

Τμηματοποίηση εικόνας με βάση την περιοχή

D. Barbosa πρότεινε μια νέα τεχνική τμηματοποίησης εικόνας η οποία συνδυάζει τις πληροφορίες που βασίζονται στις ακμές και τις περιοχές με τη φασματική μέθοδο χρησιμοποιώντας μορφολογικούς αλγορίθμους Watershed. Αρχικά χρησιμοποιείται φίλτρο θορύβου με κλίση μεγέθους σε ένα στάδιο προεπεξεργασίας, δεύτερον, γίνεται προ-τμηματοποίηση χρησιμοποιώντας συγχώνευση περιοχών, στη συνέχεια δημιουργείται γράφημα ομοιότητας περιοχών και τέλος πραγματοποιείται τμηματοποίηση χρησιμοποιώντας Multi Class Normalized Cut. Η μέθοδος συγκρίνεται με τις μεθόδους Mean Shift, MNCUT και JSEG χρησιμοποιώντας φυσικές εικόνες. Η προτεινόμενη τεχνική ξεπερνά τη μέθοδο φασματικής ομαδοποίησης. Ο Gang Chen διαπίστωσε ότι η γρήγορη εξαγωγή πληροφοριών αντικειμένου από μια δεδομένη εικόνα εξακολουθεί να αποτελεί πρόβλημα για την επεξεργασία εικόνας σε πραγματικό χρόνο. Διαπίστωσαν επίσης ότι οι μέθοδοι που βασίζονται σε περιοχές είναι επίσης χρονοβόρες και δεν δίνουν αποτελεσματική κατάτμηση. Πρότεινε μια νέα μέθοδο βασισμένη σε περιοχές που βασίζεται στη μέθοδο ελαχίστων τετραγώνων προκειμένου να ανιχνεύονται αντικείμενα με ευκρίνεια. Χρησιμοποίησε έναν πίνακα βαρών για τη μέθοδο που βασίζεται στην περιοχή, ο οποίος λαμβάνει επίσης υπόψη τις τοπικές

πληροφορίες και επίσης η χρήση της μεθόδου ελαχίστων τετραγώνων παρέχει βέλτιστη και γρήγορη τμηματοποίηση. Η σύγκριση της μεθόδου τους πραγματοποιείται με τη μέθοδο Otsu και τη μέθοδο Chan-Vese χρησιμοποιώντας την εικόνα Lena. Η μεθόδός τους μπορεί να εξάγει τα χαρακτηριστικά με μεγαλύτερη ακρίβεια από άλλες μεθόδους. Οι Zhen Hua, Yewei Li πρότειναν μια νέα μέθοδο τμηματοποίησης εικόνας που βασίζεται στη βελτιωμένη οπτική προσοχή και στην προσέγγιση ανάπτυξης περιοχών. Οι τιμές του γκρι και οι ακμές της εικόνας εισόδου εξάγονται χρησιμοποιώντας φίλτρο Gabor και φίλτρα Guass-Laplace. Στη συνέχεια χρησιμοποιούνται μέθοδοι ANN για την εξαγωγή της περιοχής ενδιαφέροντος. Πραγματοποιούνται πειράματα σε εικόνες φυσικών σκηνών και διαπιστώνεται ότι ο αλγόριθμός τους όχι μόνο τμηματοποίησε τέλεια την εικόνα αλλά και βρήκε τις άκρες ενδιαφέροντος που άλλοι δεν μπορούν. Ο Tiancan Mei ισχυρίζεται ότι το τυχαίο πεδίο Markov (MRF) πάσχει από έλλειψη χειρισμού του μεγάλου εύρους αλληλεπίδρασης. Προκειμένου να ξεπεραστεί αυτό το μειονέκτημα του MRF, παρουσίασε μια νέα μέθοδο τμηματοποίησης εικόνας με επίβλεψη, δηλαδή τη μέθοδο τμηματοποίησης πολλαπλών κλιμάκων με βάση την περιοχή. Οι εικόνες φυσικής σκηνής χρησιμοποιούνται ως σύνολο δεδομένων. Χρησιμοποιώντας την περιοχή ως παράμετρο στο μοντέλο MRF πολλαπλών κλιμάκων, οι αλγόριθμοί τους έχουν καλύτερες επιδόσεις από άλλες τεχνικές. Από τα αποτελέσματα παρατηρείται ότι η RSMAP βελτιώνει τη μέθοδο MSAP που χρησιμοποιείται για την κατάτμηση εικόνας.

ΔΙΕΞΑΓΩΓΗ ΠΕΙΡΑΜΑΤΟΣ

Για τον σκοπό του εντοπισμού και ανίχνευσης μεδουσών επικίνδυνο στις ελληνικές θάλασσες θα έπρεπε να προβούμε σε μελέτες στο φυσικό περιβάλλον. Πιο συγκεκριμένα, χρησιμοποιώντας υποβρύχιες συσκευές λάβαμε εικόνες και βίντεο από το λιμάνι της κασιόπης. Έπειτα, χρησιμοποιώντας τις τεχνικές εντοπισμού που έχουμε αναφέρει και αναλύσει εκτενώς τις προηγούμενες σελίδες, προβήκαμε στις απαραίτητες εκείνες ενέργειες ώστε να οδηγηθούμε σε όσο το δυνατόν ασφαλέστερα συμπεράσματα. Πάμε να δούμε τα βήματα τα οποία ακολουθήσαμε για την διεξαγωγή της συγκεκριμένης εργαστηριακής μελέτης.

Αρχικά ξεκινήσαμε να ασχολούμεθα με τα νευρωνικά δίκτυα. Με βάση το υλικό το οποίο λάβαμε από τις υποβρύχιες συσκευές φτιάξαμε 2 σύνολα. Το πρώτο σύνολο το πρώτο αποτελείται από εικόνες οι οποίες εμπεριέχουν μέδουσες. Το δεύτερο σετ ή σύνολο αποτελείται από εικόνες διαφορετικής φύσεως σε σχέση με τις εικόνες του πρώτου συνόλου. Στη συνέχεια χρησιμοποιώντας το εργαλείο google colab, προβήκαμε με τη χρήση των κατάλληλων κατάλληλων γραμμών κώδικα Στην ταξινόμηση των εικόνων, συμπεριλαμβανομένων και των 2 συνόλων. Για να επιτύχουμε το σκοπό μας δηλαδή την ταξινόμηση σε κάθε φάκελο δημιουργήσαμε

υποφακέλους εκμάθησης και δοκιμής. Έπειτα αξιοποιώντας τους κώδικες ανοιχτού λογισμικού υλοποιήσαμε την κατάλληλη επεξεργασία και μορφοποίηση των δειγμάτων μας δηλαδή των εικόνων μας. Ως αποτέλεσμα , χρησιμοποιώντας την αρχιτεκτονική νευρωνικών δικτύων υλοποιήσαμε επιτυχημένη ταξινόμηση δειγμάτων. Αυτό αποτυπώνεται στις παρακάτω εικόνες:

26/34

25/3/24, 8:44 μμ.

NEURAL NETWORK - Colaboratory

(1.0000000000000002, 0.0784313725490196)

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X = scaler.fit_transform(X)

X[0].max(), X[0].min()

(1.0000000000000002, 0.0784313725490196)

X[1]

array([0.0745998 , 0.07843137, 0.08235294, ..., 0.0627451 , 0.0627451 ,
0.0627451])

TRAIN AND TEST SET

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 1)

X_train.shape, y_train.shape

((101, 16384), (101,))

X_test.shape, y_test.shape

((26, 16384), (26,))

BUILDING AND TRAINING THE NEURAL NETWORK

(16384 + 2) / 2

8193.0

https://colab.research.google.com/drive/158mjEH4ZNPQgBx9cIUrQmPdOq9CatP#scrollTo=KLJV17TC7Zxn&printMode=true

27/34

25/3/24, 8:44 μμ.

NEURAL NETWORK - Colaboratory

network1 = tf.keras.models.Sequential()
network1.add(tf.keras.layers.Dense(input_shape=(16384,), units=8193, activation='relu'))
network1.add(tf.keras.layers.Dense(units=8193, activation = 'relu'))
network1.add(tf.keras.layers.Dense(units = 1, activation = 'sigmoid'))

network1.summary()

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 8193)	134242305
dense_1 (Dense)	(None, 8193)	67133442
dense_2 (Dense)	(None, 1)	8194

Total params: 201383941 (768.22 MB)
Trainable params: 201383941 (768.22 MB)
Non-trainable params: 0 (0.00 Byte)

network1.compile(optimizer='Adam', loss='binary_crossentropy', metrics = ['accuracy'])

history = network1.fit(X_train, y_train, epochs=50)

Epoch 1/50
4/4 [=====] - 3s 69ms/step - loss: 131.5071 - accuracy: 0.5545
Epoch 2/50
4/4 [=====] - 0s 55ms/step - loss: 7.8565 - accuracy: 0.4752
Epoch 3/50
4/4 [=====] - 0s 55ms/step - loss: 10.7009 - accuracy: 0.6832
Epoch 4/50
4/4 [=====] - 0s 56ms/step - loss: 9.7116 - accuracy: 0.3861
Epoch 5/50
4/4 [=====] - 0s 55ms/step - loss: 6.9558 - accuracy: 0.7426
Epoch 6/50
4/4 [=====] - 0s 56ms/step - loss: 3.7504 - accuracy: 0.3267
Epoch 7/50
4/4 [=====] - 0s 55ms/step - loss: 2.3446 - accuracy: 0.7228
Epoch 8/50
4/4 [=====] - 0s 56ms/step - loss: 1.8458 - accuracy: 0.7426
Epoch 9/50
4/4 [=====] - 0s 56ms/step - loss: 0.7693 - accuracy: 0.5842
Epoch 10/50
4/4 [=====] - 0s 56ms/step - loss: 0.6555 - accuracy: 0.6733
Epoch 11/50
4/4 [=====] - 0s 56ms/step - loss: 0.4041 - accuracy: 0.8218
Epoch 12/50
4/4 [=====] - 0s 55ms/step - loss: 0.3821 - accuracy: 0.8119

https://colab.research.google.com/drive/158mjEH4ZNPQgBx9cIUrQmPdOq9CatP#scrollTo=KLJV17TC7Zxn&printMode=true

28/34

```

Epoch 13/50
4/4 [=====] - 0s 55ms/step - loss: 0.3358 - accuracy: 0.8614
Epoch 14/50
4/4 [=====] - 0s 56ms/step - loss: 0.3194 - accuracy: 0.8317
Epoch 15/50
4/4 [=====] - 0s 56ms/step - loss: 0.3208 - accuracy: 0.8713
Epoch 16/50
4/4 [=====] - 0s 56ms/step - loss: 0.3066 - accuracy: 0.8515
Epoch 17/50
4/4 [=====] - 0s 56ms/step - loss: 0.3400 - accuracy: 0.8317
Epoch 18/50
4/4 [=====] - 0s 56ms/step - loss: 0.2903 - accuracy: 0.9010
Epoch 19/50
4/4 [=====] - 0s 56ms/step - loss: 0.2527 - accuracy: 0.9010
Epoch 20/50
4/4 [=====] - 0s 56ms/step - loss: 0.2578 - accuracy: 0.8911
Epoch 21/50
4/4 [=====] - 0s 56ms/step - loss: 0.1783 - accuracy: 0.9208
Epoch 22/50
4/4 [=====] - 0s 57ms/step - loss: 0.1865 - accuracy: 0.9208
Epoch 23/50
4/4 [=====] - 0s 56ms/step - loss: 0.4002 - accuracy: 0.7921
Epoch 24/50
4/4 [=====] - 0s 56ms/step - loss: 0.2467 - accuracy: 0.8614
Epoch 25/50
4/4 [=====] - 0s 55ms/step - loss: 0.3035 - accuracy: 0.8317
Epoch 26/50
4/4 [=====] - 0s 55ms/step - loss: 0.1685 - accuracy: 0.9109
Epoch 27/50
4/4 [=====] - 0s 55ms/step - loss: 0.1531 - accuracy: 0.9109
Epoch 28/50
4/4 [=====] - 0s 56ms/step - loss: 0.1126 - accuracy: 0.9406
Epoch 29/50
4/4 [=====] - 0s 55ms/step - loss: 0.1450 - accuracy: 0.9307

```

EVALUATING THE NEURAL NETWORK

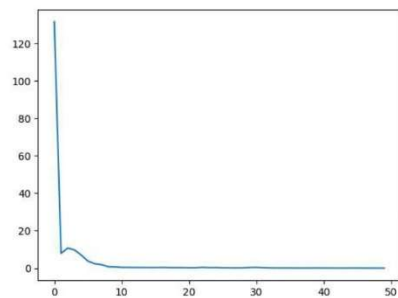
```

history.history.keys()

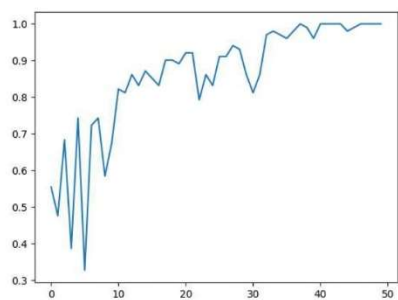
dict_keys(['loss', 'accuracy'])

plt.plot(history.history['loss']);

```



```
plt.plot(history.history['accuracy']);
```



X_test.shape

NEURAL NETWORK - Colaboratory

```

predictions = network1.predict(X_test)

print(predictions)

1/1 [.....] - 0s 18ms/step
array([[ 0.90000000e+00],
       [ 9.3433820e-01],
       [ 9.9990249e-01],
       [ 9.9871588e-01],
       [ 8.3065940e-01],
       [ 9.9991572e-01],
       [ 9.9624652e-01],
       [ 3.3417681e-01],
       [ 9.9999898e-01],
       [ 6.2973334e-01],
       [ 1.00000000e+00],
       [ 9.9986053e-01],
       [ 6.0768740e-01],
       [ 9.9964720e-01],
       [ 9.9608684e-01],
       [ 6.4314365e-01],
       [ 9.9924809e-01],
       [ 9.7874874e-01],
       [ 1.9769319e-02],
       [ 2.0121528e-04],
       [ 9.9997926e-01],
       [ 9.8829532e-01],
       [ 8.8604450e-01],
       [ 9.9542144e-01],
       [ 9.9944900e-01],
       [ 1.00000000e+00]], dtype=float32)

```

[illegible]

31/34

NEURAL NETWORK - Colaboratory

y_test

[illegible]

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test, predictions)

0.9230769230769231
```

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, predictions)
cm
```

```
array([[ 1,  0],
       [ 2, 23]])
```

```
from sklearn.metrics import classification_report
print(classification_report(y_test, predictions))
```

	precision	recall	f1-score	support
0	0.33	1.00	0.50	1
1	1.00	0.92	0.96	25
accuracy			0.92	26
macro avg	0.67	0.96	0.73	26
weighted avg	0.97	0.92	0.94	26

SAVING AND LOADING THE NETWORK

```
model_json = network1.to_json()
with open('network1.json', 'w') as json_file:
    json_file.write(model_json)
```

32/34

25/3/24, 8:44 μ.μ.

NEURAL NETWORK - Colaboratory

```

from keras.models import save_model
network1_saved = save_model(network1, 'weights1.hdf5')

<ipython-input-45-2d2cfff364421>:2: UserWarning: You are saving your model as an HDF5 file via 'model.save()'. This file format is considered legacy. We recommend using instead
network1_saved = save_model(network1, 'weights1.hdf5')

with open('network1.json') as json_file:
    json_saved_model = json_file.read()
json_saved_model

{
  "class_name": "Sequential",
  "config": {
    "name": "sequential",
    "layers": [
      {
        "module": "keras.layers",
        "class_name": "InputLayer",
        "config": {
          "batch_input_shape": [null, 16384],
          "dtype": "float32",
          "sparse": false,
          "ragged": false,
          "name": "dense_input",
          "registered_name": null,
          "module": "keras.layers",
          "class_name": "Dense",
          "config": {
            "name": "dense",
            "trainable": true,
            "dtype": "float32",
            "batch_input_shape": [null, 16384],
            "units": 8193,
            "activation": "relu",
            "use_bias": true,
            "kernel_initializer": {
              "module": "keras.initializers",
              "class_name": "GlorotUniform",
              "config": {
                "seed": null,
                "registered_name": null,
                "bias_initializer": {
                  "module": "keras.initializers",
                  "class_name": "Zeros",
                  "config": {},
                  "registered_name": null,
                  "kernel_regularizer": {
                    "module": "keras.regularizers",
                    "class_name": "L2",
                    "config": {
                      "kernel_constraint": null,
                      "bias_constraint": null,
                      "name": "l2_regularizer"
                    }
                  }
                }
              }
            }
          }
        }
      ]
    }
  }
}

network1_loaded = tf.keras.models.model_from_json(json_saved_model)
network1_loaded.load_weights('/content/weights1.hdf5')
network1_loaded.compile(loss = 'binary_crossentropy', optimizer = 'Adam', metrics = ['accuracy'])

network1_loaded.summary()

Model: "sequential"
-----
Layer (type)                Output Shape              Param #
-----
dense (Dense)                (None, 8193)              134242385
dense_1 (Dense)              (None, 8193)              67133442
dense_2 (Dense)              (None, 1)                 8194
-----
Total params: 201383941 (768.22 MB)
Trainable params: 201383941 (768.22 MB)
Non-trainable params: 0 (0.00 Byte)

```

CLASSIFYING ONE SINGLE IMAGE

```

X_test[25].shape
(16384,)

```

<https://colab.research.google.com/drive/158mjEH4ZNPQgBx9cUjQmPdQo9CaiP#scrollTo=KLJV17CTZx&printMode=true>

33/34

25/3/24, 8:44 μ.μ.

NEURAL NETWORK - Colaboratory

```

test_image
array([[18., 16., 11., ..., 7., 8., 7.]])

test_image = X_test[25]
test_image = scaler.inverse_transform(test_image.reshape(1, -1))

test_image
array([[33., 21., 30., ..., 7., 6., 6.]])

cv2.imshow(test_image.reshape(width, height))

network1_loaded.predict(test_image)[0][0]
1/1 [=====] - 0s 16ms/step
1.0

if network1_loaded.predict(test_image)[0][0] < 0.5:
    print('FISH')
else:
    print('JELLY')

1/1 [=====] - 0s 17ms/step
JELLY

```

<https://colab.research.google.com/drive/158mjEH4ZNPQgBx9cUjQmPdQo9CaiP#scrollTo=KLJV17CTZx&printMode=true>

34/34

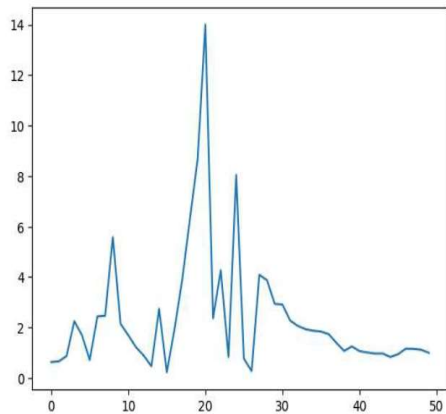
Στην επόμενη προσέγγιση, αυτή των CNN πέραν της ταξινόμησης εικόνων, δοκιμάσαμε και την αποτελεσματικότητα του μοντέλου μας σε άγνωστο αντικείμενο. Πιο συγκεκριμένα, έχοντας τραβήξει βίντεο με αλλά και χωρίς μέδουσα, ελέγχουμε την αξιοπιστία του προηγούμενου μοντέλου. Όπως φαίνεται και στις παρακάτω

εικόνες, το μοντέλο μας δεν έχει υψηλό ποσοστό ακριβείας.

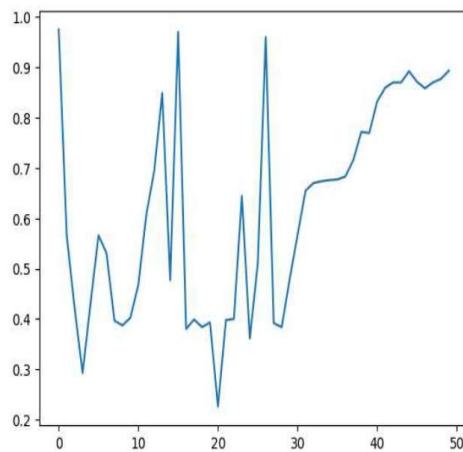
```
19/3/24, 10:14 π.μ. convolutional neural network 2 - Colaboratory
Epoch 26/50
18/18 [=====] - 7s 404ms/step - loss: 0.2324 - accuracy: 0.9712 - val_loss: 0.7674 - val_accuracy: 0.5099
Epoch 27/50
18/18 [=====] - 6s 342ms/step - loss: 0.0263 - accuracy: 0.9928 - val_loss: 0.2729 - val_accuracy: 0.9602
Epoch 28/50
```

```
history.history.keys()
dict_keys(['loss', 'accuracy', 'val_loss', 'val_accuracy'])
```

```
plt.plot(history.history['val_loss']);
```



```
plt.plot(history.history['val_accuracy']);
```



```
test_dataset.class_indices
```

```
{'blue': 0, 'jelly': 1}
```

```
predictions = network.predict(test_dataset)
predictions
```

```
855/855 [=====] - 5s 6ms/step
array([[1.0000000e+00, 1.7159055e-13],
       [1.0000000e+00, 1.4417109e-13],
       [1.0000000e+00, 9.4925187e-14],
       ...,
       [8.9362439e-18, 1.0000000e+00],
       [5.2927722e-18, 1.0000000e+00],
       [1.6126233e-17, 1.0000000e+00]], dtype=float32)
```

```
predictions = np.argmax(predictions, axis = 1)
predictions
```

[illegible]

convolutional neural network 2 - Colaboratory

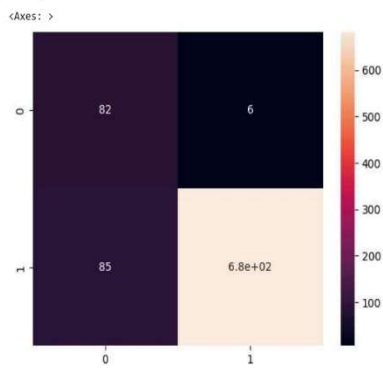
```
from sklearn.metrics import accuracy_score
accuracy_score(test_dataset.classes, predictions)

0.8935672514619883

test_dataset.class_indices

{'blue': 0, 'jelly': 1}

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(test_dataset.classes, predictions)
sns.heatmap(cm, annot=True)
```



```
from sklearn.metrics import classification_report
print(classification_report(test_dataset.classes, predictions))
```

```

              precision    recall  f1-score   support

     0       0.49      0.93   0.64         88
     1       0.99      0.89   0.94        767

 accuracy          0.74      0.91   0.89        855
 macro avg          0.74      0.91   0.79        855
 weighted avg          0.94      0.89   0.91        855

```

```
model_json = network.to_json()
with open('network.json', 'w') as json_file:
    json_file.write(model_json)
```

```
from keras.models import save_model
network_saved = save_model(network, '/content/weights.hdf5')
```

```
<ipython-input-27-32bf80f9a744>:2: UserWarning: You are saving your model as an HDF5 file via 'model.save()'. This file format is considered legacy. We recommend using instead
network_saved = save_model(network, '/content/weights.hdf5')
```

```
with open('network.json', 'r') as json_file:
    json_saved_model = json_file.read()
json_saved_model

{'class_name': 'Sequential', 'config': {'name': 'sequential', 'layers': [{'module': 'keras.layers', 'class_name': 'InputLayer', 'config': {'batch_input_shape': [null, 128, 128, 3], 'dtype': 'float32', 'sparse': false, 'ragged': false, 'name': 'conv2d_input'}, 'registered_name': null}, {'module': 'keras.layers', 'class_name': 'Conv2D', 'config': {'name': 'conv2d', 'trainable': true, 'dtype': 'float32', 'batch_input_shape': [null, 128, 128, 3], 'filters': 32, 'kernel_size': [3, 3], 'strides': [1, 1], 'padding': 'valid', 'data_format': 'channels_last', 'dilation_rate': [1, 1], 'group': 1, 'activation': 'relu', 'use_bias': true, 'kernel_initializer': {'module': 'keras.initializers', 'class_name': 'GlorotUniform', 'config': {'seed': null}, 'registered_name': null}, 'bias_initializer': {'module': 'keras.initializer
```

```
network_loaded = tf.keras.models.model_from_json(json_saved_model)
network_loaded.load_weights('weights.hdf5')
network_loaded.compile(loss = 'categorical_crossentropy', optimizer='Adam', metrics=['accuracy'])
```

```
network_loaded.summary()
```

Model: "sequential"		
Layer (type)	Output Shape	Param #
=====		
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 32)	0
conv2d_2 (Conv2D)	(None, 28, 28, 32)	9248
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 32)	0
flatten (Flatten)	(None, 6272)	0
dense (Dense)	(None, 3137)	19678401
dense_1 (Dense)	(None, 3137)	9843906
dense_2 (Dense)	(None, 2)	6276
=====		
Total params: 29547975 (112.72 MB)		
Trainable params: 29547975 (112.72 MB)		
Non-trainable params: 0 (0.00 Byte)		

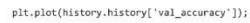
```
image = cv2.imread('/content/jelly and other/test.set/jelly/jellyfish1003.bmp')
```

```
cv2.imshow(image)
```



```
image = cv2.resize(image, (128, 128))  
cv2.imshow(image)
```

Έχοντας προς επίλυση το ζήτημα του χαμηλού σκορ, αποφασίσαμε να ακολουθήσουμε μια διαφορετική πρόσεγγιση. Ελλείπει πολλών εικονοστοιχείων για την εκπαίδευση του μοντέλου μας, προχωρήσαμε σε μετατροπή του βίντεο σε frames. Λάβαμε περί τις 1100 εικόνες. Τοποθετώντας έναν ορισμένο αριθμό στο σύνολο δοκιμής και έναν άλλο αριθμό δειγμάτων στο σύνολο εκμάθησης, ξανατρέξαμε τον κώδικα. Τα αποτελέσματα ήταν τα ακόλουθα:



7/16

https://colab.research.google.com/drive/12-aqm_-t4XGQf8QDvgvD5so86CfZNJqY#printMode=true


```
19/3/24, 10:15 π.μ. Convolutional Neural Network Jellyfish.ipynb - Colaboratory

<ipython-input-25-32bf80f9a744>:2: UserWarning: You are saving your model as an HDF5 file via 'model.save()'. This file format is considered legacy. We recommend using instead
network_saved = save_model(network, '/content/weights.hdf5')

with open('network.json', 'r') as json_file:
    json_saved_model = json_file.read()
    json_saved_model

{'class_name': 'Sequential', 'config': {'name': 'sequential', 'layers': [{'module': 'keras.layers', 'class_name': 'InputLayer', 'config': {'batch_input_shape': [null, 128, 128, 3], 'dtype': 'float32', 'sparse': false, 'ragged': false, 'name': 'conv2d_input', 'registered_name': null}, {'module': 'keras.layers', 'class_name': 'Conv2D', 'config': {'name': 'conv2d', 'trainable': true, 'dtype': 'float32', 'batch_input_shape': [null, 128, 128, 3], 'filters': 32, 'kernel_size': [3, 3], 'strides': [1, 1], 'padding': 'valid', 'data_format': 'channels_last', 'dilation_rate': [1, 1], 'group': 1, 'activation': 'relu', 'use_bias': true, 'kernel_initializer': {'module': 'keras.initializers', 'class_name': 'GlorotUniform', 'config': {'seed': null}, 'registered_name': null}, 'bias_initializer': {'module': 'keras.initializers', 'config': {'seed': null}, 'registered_name': null}}]}]}

network_loaded = tf.keras.models.model_from_json(json_saved_model)
network_loaded.load_weights('weights.hdf5')
network_loaded.compile(loss = 'categorical_crossentropy', optimizer='Adam', metrics=['accuracy'])

network_loaded.summary()

Model: "sequential"
Layer (type) Output Shape Param #
-----
conv2d (Conv2D) (None, 126, 126, 32) 896
max_pooling2d (MaxPooling2D) (None, 63, 63, 32) 0
conv2d_1 (Conv2D) (None, 61, 61, 32) 9248
max_pooling2d_1 (MaxPooling2D) (None, 30, 30, 32) 0
conv2d_2 (Conv2D) (None, 28, 28, 32) 9248
max_pooling2d_2 (MaxPooling2D) (None, 14, 14, 32) 0
flatten (Flatten) (None, 6272) 0
dense (Dense) (None, 3137) 19678401
dense_1 (Dense) (None, 3137) 9843906
dense_2 (Dense) (None, 2) 6276
Total params: 29547975 (112.72 MB)
Trainable params: 29547975 (112.72 MB)

https://colab.research.google.com/drive/12-aqm_-14XGQf8QDvgvD5so86CfZJqY#printMode=true 11/16
```

```
19/3/24, 10:15 π.μ. Convolutional Neural Network Jellyfish.ipynb - Colaboratory

Non-trainable params: 0 (0.00 Byte)

image = cv2.imread('/content/jellyfish and blue/test set/jellyfish/jellyfish1003.bmp')
cv2.imshow(image)

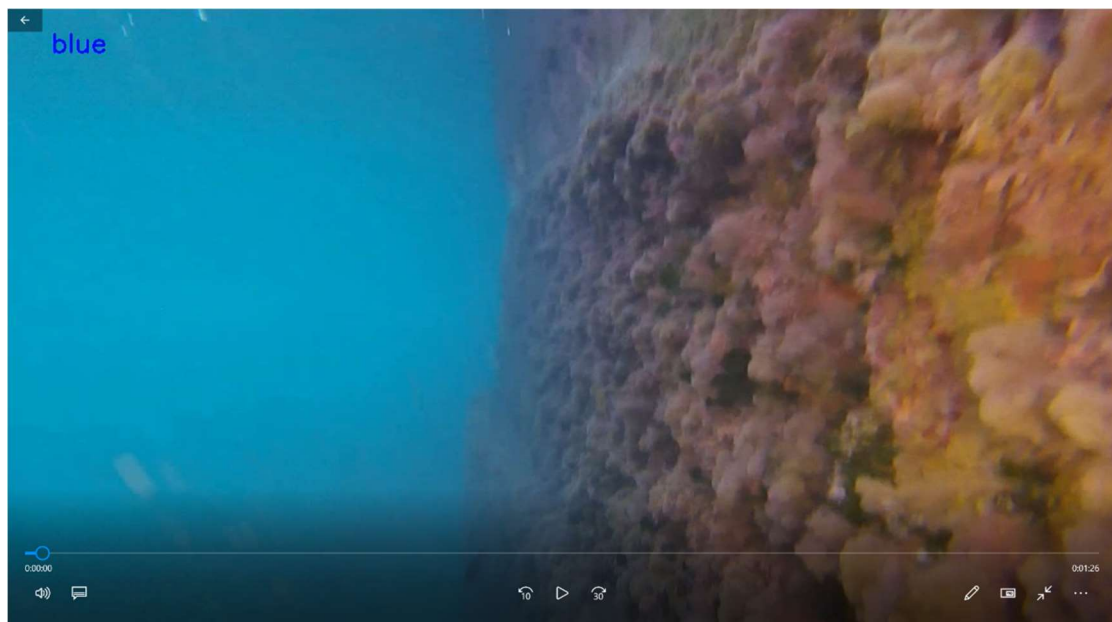
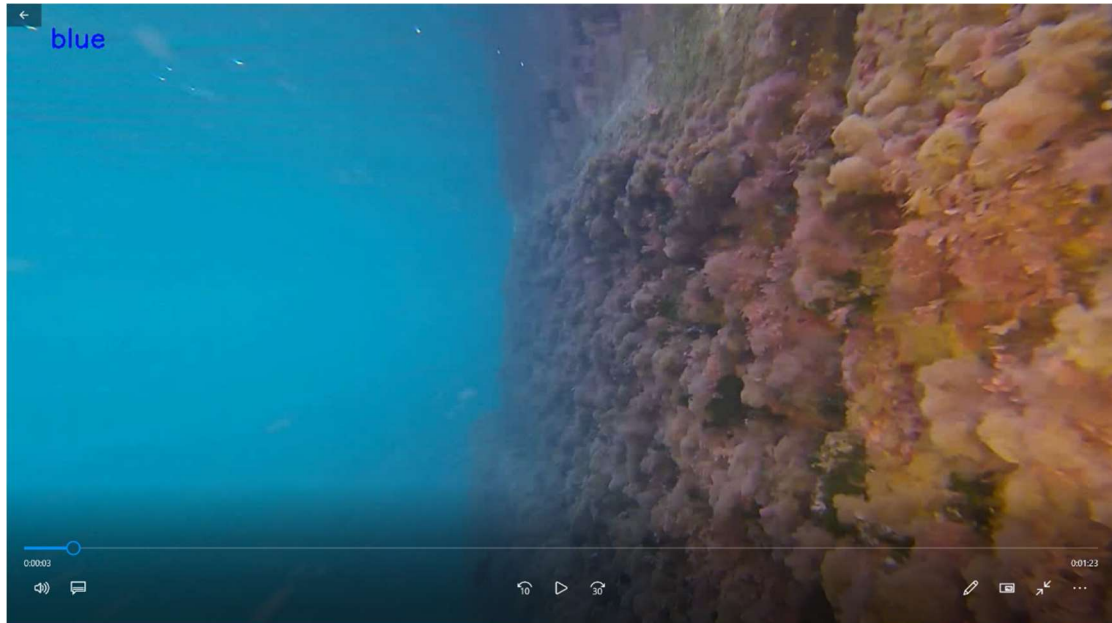


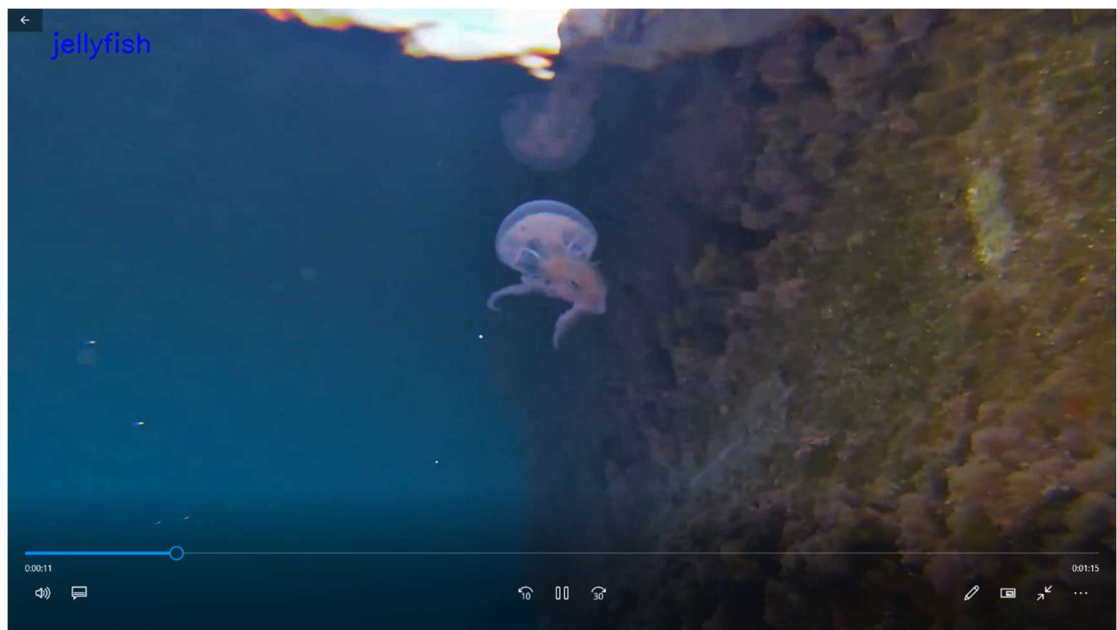
image = cv2.resize(image, (128, 128))
cv2.imshow(image)

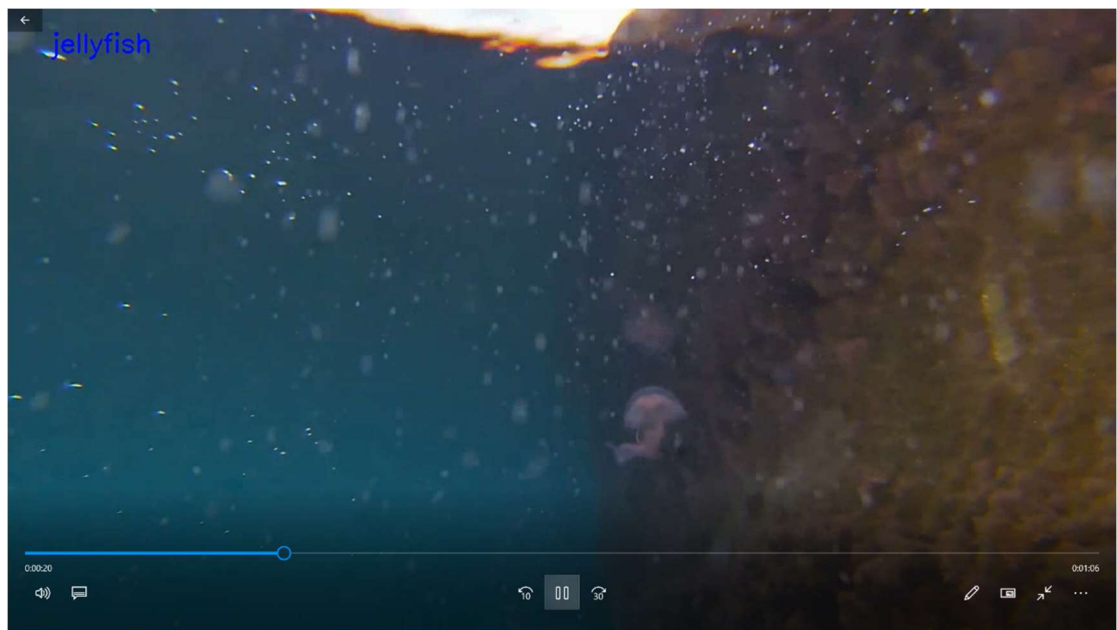
https://colab.research.google.com/drive/12-aqm_-14XGQf8QDvgvD5so86CfZJqY#printMode=true 12/16
```

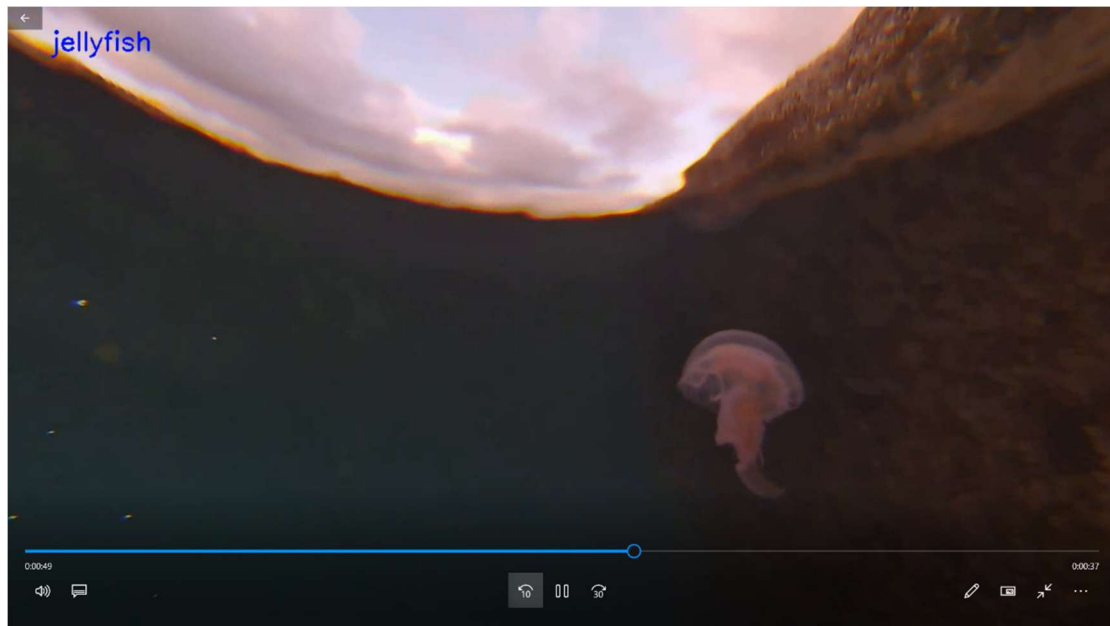
Εν κατακλείδι, το f1 score ανέβηκε από 0,64 σε 0,99 για το αντικείμενο 0, ενώ για το

αντικείμενο 1 δεν υπήρξε ουσιώδης διαφοροποίηση στο σκορ. Αυτό συνέβη διότι το μοντέλο μας στην πρώτη περίπτωση είχε μεγάλο ποσοστό αναγνώρισης ψευδο-θετικές απεικονίσεις του αντικειμένου 0. Παρακάτω παραθέτουμε λίγα στιγμιότυπα από το βίντεο







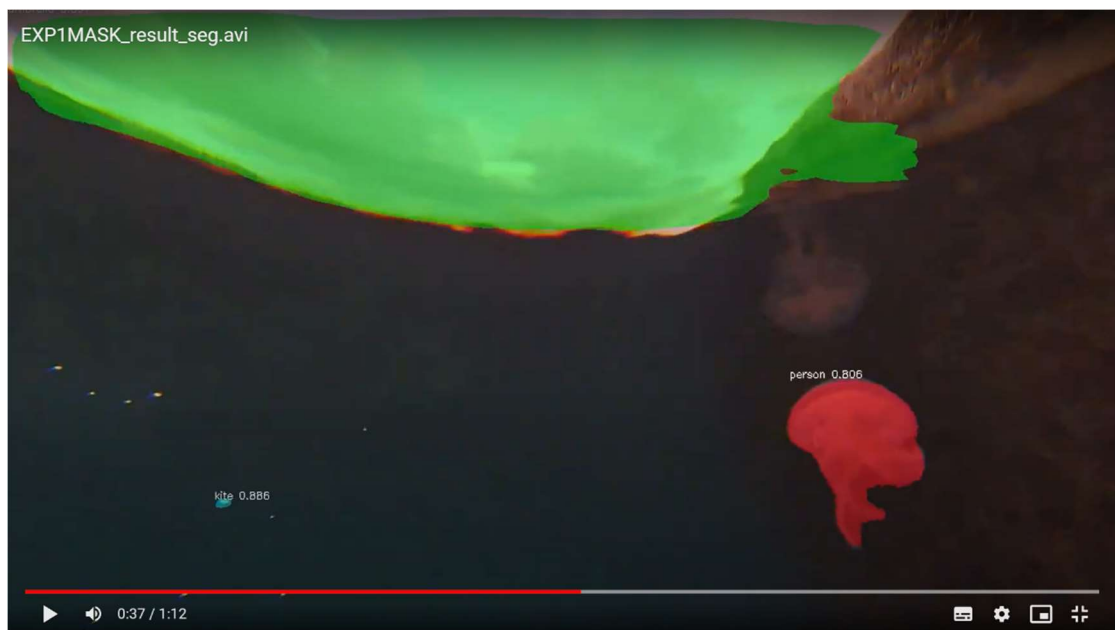


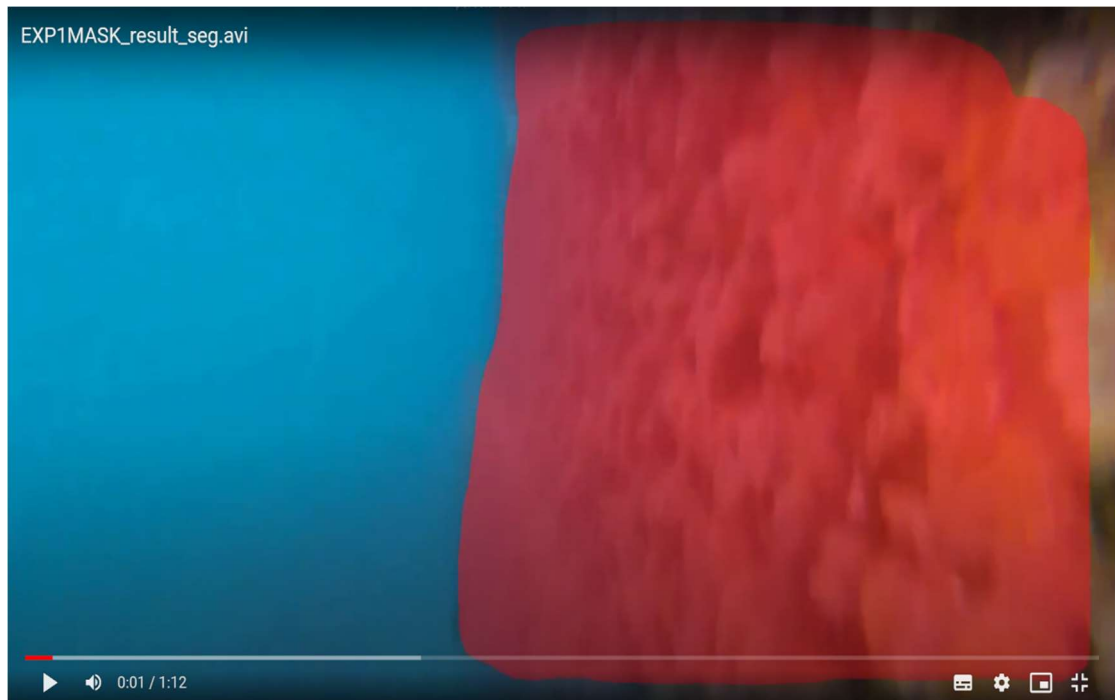
Οι επόμενες δύο προσεγγίσεις είναι πανομοιότυπες. Είναι και οι δύο αρχιτεκτονικές Συνελκτικών Δικτύων, διαφέρουν όμως στον τρόπο που αναπτύχθηκαν. Και στις δύο αυτές προσεγγίσεις χρησιμοποιούμε το ίδιο βίντεο με την προηγούμενη αρχιτεκτονική για να ορίσουμε τον στόχο ανίχνευσης. Και οι δύο αυτές αρχιτεκτονικές χρησιμοποιούν μια συγκεκριμένη βάση εκπαιδευμένων αντικειμένων, ώστε να ταυτοποιήσουν τον στόχο. Στην προσέγγιση YOLO σχηματίζεται ένα οριοθετημένο τετράγωνο γύρω από το αντικείμενο ενδιαφέροντος ενώ στο MASK-R-CNN σχηματίζεται και μια "μάσκα" γύρω από το αντικείμενο. Το πλεονέκτημα της μάσκας έγκειται στην διαφοροποίηση ενός αντικειμένου από άλλα αντικείμενα της ίδιας φύσης ή διαφορετικά. Ο μείζον λόγος που δεν επιλέξαμε να αναπτύξουμε στην συγκεκριμένη εργασία αυτές τις δυο αρχιτεκτονικές μεμονωμένα έγκειται στην γηγενή τους αρχιτεκτονική. Για να μπορούμε να εντοπίζουμε τις μωβ μέδουσες χρησιμοποιώντας αυτές τις δύο τεχνικές θα πρέπει να εκπαιδύσουμε τα νευρωνικά δίκτυα ακριβώς όπως στα Συνελκτικά Νευρωνικά

Δίκτυα. Ο βασικός λόγος που απαιτείται αυτή η διαδικασία έγκειται ότι η μέδουσα δεν ανήκει στην βάση εκπαιδευμένων αντικειμένων(coco). Κατ επέκταση, το κυρίαρχο βήμα στη ανίχνευση βρίσκεται στην εκπαίδευση των νευρωνικών δικτύων. Παρακάτω παραθέτουμε στιγμιότυπα των βίντεο από τις δύο τελευταίες αρχιτεκτονικές.

MASK-R-CNN

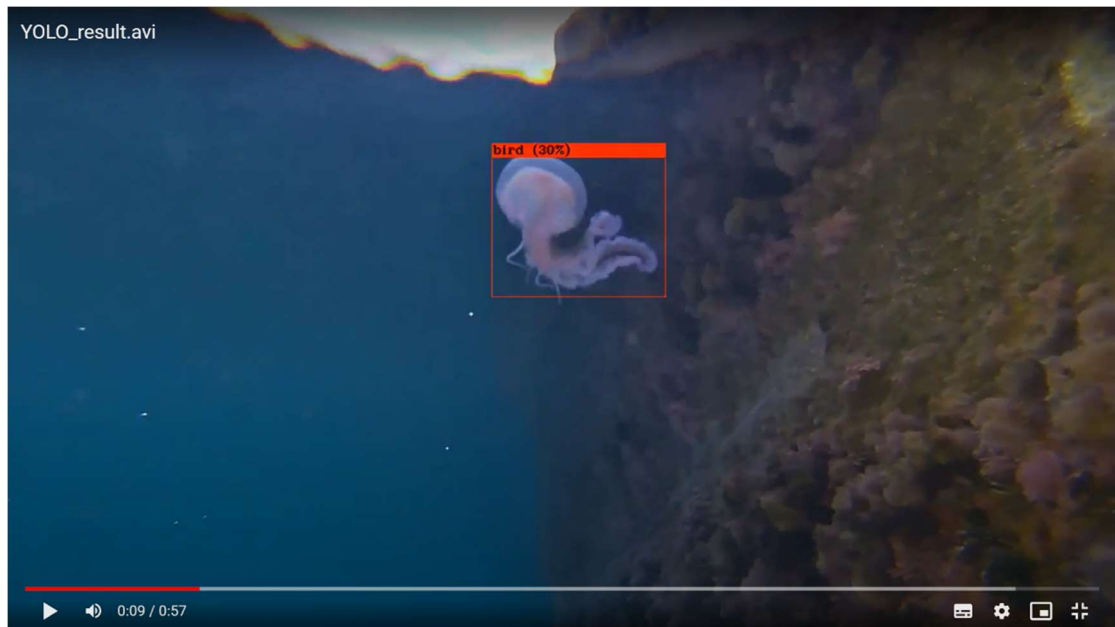






Όπως βλέπουμε, ελέω των κινήσεων της μέδουσας, σχηματίζονται διάφορα σχήματα. Αυτά τα σχήματα γίνονται αντιληπτά από τον αλγόριθμο ως κάποια από τα αντικείμενα που υπάρχουν στην εκπαιδευμένη του βάση.

YOLO







ΚΑΤΑΚΛΕΙΔΑ

Έχοντας αναλύσει εκτενώς τις διάφορες αρχιτεκτονικές των νευρωνικών δικτύων μπορούμε να καταλήξουμε και να οδηγηθούμε σε συγκεκριμένα συμπεράσματα. Πρώτον, η διαδικασία συλλογής δεκάδων χιλιάδων εικονοστοιχείων το αντικείμενο της επιλογής μας (μωβ μέδουσες) είναι ζωτικής σημασίας για τον ακριβή εντοπισμό. Δεύτερον, οι απαιτήσεις σε επεξεργαστική ισχύ για την ανάπτυξη αυτών των αρχιτεκτονικών είναι προαπαιτούμενες ώστε να έχουμε άμεσα και ορατά αποτελέσματα. Τρίτον, η Ικανότητα σχεδιασμού και ανάπτυξης περιπλοκών συν ελεγκτικών νευρωνικών δικτύων, Αποτελεί τον καθοριστικό παράγοντα επιτυχίας του έργου σε βάθος χρόνου. Πάνω σε αυτήν την αρχιτεκτονική μπορούμε να αναπτύξουμε και να δοκιμάσουμε διάφορους παραμέτρους Όστε να βελτιώσουμε την αποτελεσματικότητα το αλγόριθμου. Τέλος η ύπαρξη συγκεκριμένου τύπου υποβρύχιων συσκευών καθώς και το κτίσιμο τον γεφυρών επικοινωνίας ανάμεσα σε αυτές τις συσκευές και τα τερματικά των υπολογιστών στα γραφεία των προγραμματιστών, μπορούν να οδηγήσουν σε αρμονική συμβίωση την ανθρωπότητα με τα διάφορα ξενικά είδη τα οποία ελέω της κλιματικής αλλαγής θα εμφανίζονται όλο και πιο συχνά στις ακτές της Μεσογείου και όχι μόνο.

ΒΙΒΛΙΟΓΡΑΦΙΑ:

[1] M. Yasmin, S. Mohsin, I. Irum, and M. Sharif, "Content based image retrieval by shape, color and relevance feedback," Life Science Journal, vol. 10, 2013.

- [2] M. Rehman, M. Iqbal, M. Sharif, and M. Raza, "Content based image retrieval: survey," *World Applied Sciences Journal*, vol. 19, pp. 404-412, 2012.
- [3] M. M. S. J. Preetha, L. P. Suresh, and M. Bosco, "Image segmentation using seeded region growing," in *Proc. International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*, pp. 576-583, 2012.
- [4] W. X. Kang, Q. Q. Yang, and R. P. Liang, "The comparative research on image segmentation algorithms," in *Proc. First International Workshop on Education Technology and Computer Science*, 2009. ETCS'09. pp. 703-707, 2009
- [5] M. Yasmin, S. Mohsin, M. Sharif, M. Raza, and S. Masood, "Brain Image Analysis: A Survey," *World Applied Sciences Journal*, vol. 19, pp. 1484-1494, 2012.
- [6] S. Masood, M. Sharif, M. Yasmin, M. Raza, and S. Mohsin, "Brain Image Compression: A Brief Survey," *Research Journal of Applied Sciences*, vol. 5, 2013.
- [7] M. Yasmin, M. Sharif, S. Masood, M. Raza, and S. Mohsin, "Brain image reconstruction: A short survey," *World Applied Sciences Journal*, vol. 19, pp. 52-62, 2012.
- [8] M. Raza, M. Sharif, M. Yasmin, S. Masood, and S. Mohsin, "Brain image representation and rendering: A survey," *Research Journal of Applied Sciences*, vol. 4, 2012.
- [9] J. S. M. Sharif, S. Mohsin, and M. Raza, "Sub-holistic hidden markov model for face recognition," *Research Journal of Recent Sciences*, vol. 2, pp. 10-14, 2013.
- [10] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection" [J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [11] Chengji Liu, Yufan Tao, Jiawei Liang, Kai Li, Yihang Chen, "Object Detection Based on YOLO Network" 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC 2018)
- [12] Wenbo Lan, Jianwu Dang, Yangping Wang and Song Wang, "Pedestrian Detection Based on YOLO Network Model" 978-1-5386-6075-1/18/\$31.00 ©2018 IEEE
- [13] Joseph Redmond, Ali Farhadi, "YOLOv3: An Incremental Improvement", University of Washington
- [14] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, Jan. 2015
- [15] Scherer D, Müller A, Behnke S (2010) Evaluation of pooling operations in convolutional architectures for object recognition. In: *Artificial Neural Networks--ICANN 2010*. Springer, pp 92–101

- [16]Sermanet P, Chintala S, Lecun Y (2012) Convolutional neural networks applied to house numbers digit classification. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba. IEEE, pp 3288–3291
- [17]Sharma A, Muttoo SK (2018) Spatial Image Steganalysis Based on ResNeXt. 2018 IEEE 18th Int Conf Commun Technol 1213–1216. doi: 10.1109/ICCT.2018.8600132
- [18]Simonyan K, Zisserman A (2015) VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. ICLR 75:398–406. doi: 10.2146/ajhp170251
- [19]Sun L, Jia K, Yeung D-Y, Shi BE (2015) Human action recognition using factorized spatiotemporal convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp 4597–4605
- [20]Zhou B, Khosla A, Lapedriza A, et al (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 2921–2929