Interreg Με τη συγχρηματοδότηση της ΕΥΡΩΠΑΪΚΗΣ ΕΝΩΣΗΣ

Ελλάδα – Κύπρος

MOSAIC

ΠΑΡΑΔΟΤΕΟ 3.1:

Βιβλιογραφική αναζήτηση και στάθμη των γνώσεων. Αξιολόγηση μεθοδολογιών αναγνώρισης ακμών και χρώματος. Συλλογή δεδομένων από πολυαξονικές κατεργασίες.

ΠΕ.3	**Βιβλιογραφική αναζήτηση και στάθμη των γνώσεων. Αξιολόγηση μεθοδολογιών αναγνώρισης ακμών και χρώματος. Συλλογή δεδομένων από πολυαξονικές κατεργασίες.**
Π3.1	Τεχνική Έκθεση: Στάθμη γνώσεων σε αλγορίθμους και μεθόδους αναγνώρισης εικόνας και χρώματος με επισυναπτόμενη βιβλιογραφία

3.1.1 ΕΙΣΑΓΩΓΗ

Η σύγχρονη ψηφιακή επεξεργασία εικόνας και χρώματος αποτελεί εργαλείο υψηλής σημασίας σε πολλούς επιστημονικούς και τεχνολογικούς τομείς, επιτρέποντας την ανάλυση, την αναγνώριση και την κατηγοριοποίηση ψηφιακών δεδομένων.

Μια από τις πιο θεμελιώδης διαδικασίες στην επεξεργασία εικόνας θεωρείται αυτή της αναγνώρισης ακμών, καθώς επιτρέπει τον εντοπισμό περιοχών με έντονες μεταβολές φωτεινότητας, οι οποίες συνήθως υποδεικνύουν τα όρια μεταξύ διαφορετικών αντικειμένων. Η μελέτη των κύριων μεθόδων και αλγορίθμων, όπως οι παραδοσιακοί τελεστές (Roberts, Prewitt, Sobel, LoG) και ο αλγόριθμος Canny, βοηθά στην κατανόηση βασικών εννοιών όπως η κλίση, η συνέλιξη και η χωρική δομή της εικόνας, αποτελώντας τη βάση για πιο σύνθετες εφαρμογές ανάλυσης εικόνας.

Μια δεύτερη διαδικασία, για την οποία η αναγνώριση ακμών αποτελεί συχνά προαπαιτούμενο, είναι η τμηματοποίηση εικόνας. Η μέθοδος αυτή επιτρέπει τον διαχωρισμό της εικόνας σε περιοχές με κοινά χαρακτηριστικά, καθιστώντας δυνατή την αναγνώριση και την κατηγοριοποίηση αντικειμένων. Οι κλασικές μέθοδοι τμηματοποίησης στηρίζονται σε τεχνικές κατωφλίωσης, κατάτμηση με βάση τις ακμές ή τις περιοχές, καθώς και σε μεθόδους ομαδοποίησης και γραφημάτων, ενώ οι σύγχρονες προσεγγίσεις αξιοποιούν αλγορίθμους τεχνητής νοημοσύνης, όπως τα συνελικτικά νευρωνικά δίκτυα, τα οποία προσφέρουν αυτοματοποιημένη και εξαιρετικά ακριβή ανάλυση, ακόμη και σε σύνθετες ή θορυβώδεις εικόνες. Επιπλέον, μπορούν να διαχειριστούν μεγάλο όγκο δεδομένων, αξιοποιώντας πληροφορίες από πλήθος εικόνων για πιο αξιόπιστα αποτελέσματα, υπερβαίνοντας έτσι τους περιορισμούς που παρουσιάζουν άλλες μεθοδολογίες επεξεργασίας δεδομένων.

Μέρος της αναγνώρισης εικόνας αποτελεί και η ανάλυση και η αναγνώριση του χρώματός ενός αντικειμένου που υπάρχει σε μια εικόνα. Το αντικείμενο αυτό μπορεί να έχει προέλθει από τα διάκριτα μέρη περιεχομένου της εικόνας που έχει υποστεί τμηματοποίηση. Η ανάλυση και η αναγνώριση χρώματός αποτελεί κρίσιμο στοιχείο για πλήρη κατανόηση των χαρακτηριστικών του υπ μελέτη αντικειμένου. Η χρωματική πληροφορία παρέχει σημαντικά δεδομένα, όπως τη σύσταση των υλικών ή την αποκάλυψη φθορών. Σε πολλές εφαρμογές, η συνδυαστική χρήση της τμηματοποίησης και της αναγνώρισης χρώματος επιτρέπει την εκτενή περιγραφή της δομής και των ιδιοτήτων των αντικειμένων, δημιουργώντας έτσι ακριβή ψηφιακά αρχεία και ολοκληρωμένη ανάλυση των χαρακτηριστικών τους.

Η εφαρμογή αυτών των μεθόδων αποκτά ιδιαίτερη σημασία στον τομέα της πολιτιστικής κληρονομιάς. Η αναγνώριση εικόνας και χρώματος μέσω ψηφιακής επεξεργασίας, ιδίως με την αξιοποίηση αλγορίθμων τεχνητής νοημοσύνης, διευκολύνει τη διατήρηση, την ανάδειξη και την αποκατάσταση των πολιτιστικών αγαθών, διασφαλίζοντας την προστασία, την αειφορία και τη μεταβίβασή τους στις επόμενες γενιές.

3.1.2 Αναγνώριση Ακμών

Η αναγνώριση ακμών (edge detection) αποτελεί μία από τις θεμελιώδεις διεργασίες στον τομέα της υπολογιστικής όρασης (computer vision) [1-3], καθώς παρέχει κρίσιμες πληροφορίες για μια ευρεία γκάμα επακόλουθων οπτικών διαδικασιών, όπως αναγνώριση εικόνας (image recognition), η τμηματοποίηση εικόνας (image segmentation) και άλλες. Ο εντοπισμός των ορίων των αντικειμένων ενδιαφέροντος είναι απαραίτητος για την επιτυχία των παραπάνω διεργασιών. Σε μια εικόνα, οι ακμές εμφανίζονται στα σημεία μετάβασης μεταξύ περιοχών που διαφέρουν ως προς την ένταση (φωτεινότητα), την υφή ή το χρώμα. Η ακρίβεια με την οποία εντοπίζονται οι ακμές επηρεάζει άμεσα την ποιότητα των επόμενων σταδίων επεξεργασίας εικόνας, γεγονός που καθιστά την αξιόπιστη ανίχνευσή τους ζωτικής σημασίας. Για τον λόγο αυτό, έχουν αναπτυχθεί πληθώρα μεθόδων εντοπισμού ακμών, με στόχο την επίτευξη ισορροπίας μεταξύ υπολογιστικής αποδοτικότητας, ακρίβειας και ανθεκτικότητας σε θόρυβο.

3.1.2.1 Τύποι Ακμών σε Ψηφιακές Εικόνες

Υπάρχουν διάφοροι τύποι ακμών που προκύπτουν από διαφορετικά χαρακτηριστικά μεταβολών στην ένταση φωτεινότητας μιας εικόνας [4,5]:

- **Ακμή Βήματος (Step Edge):**
  - Απότομη μεταβολή στη φωτεινότητα-ένταση της εικόνας.
  - Παράδειγμα: μια σαφή γραμμή που διαχωρίζει δύο περιοχές έντονης αντίθεσης, όπου η φωτεινότητα αλλάζει ξαφνικά από μαύρο σε λευκό.

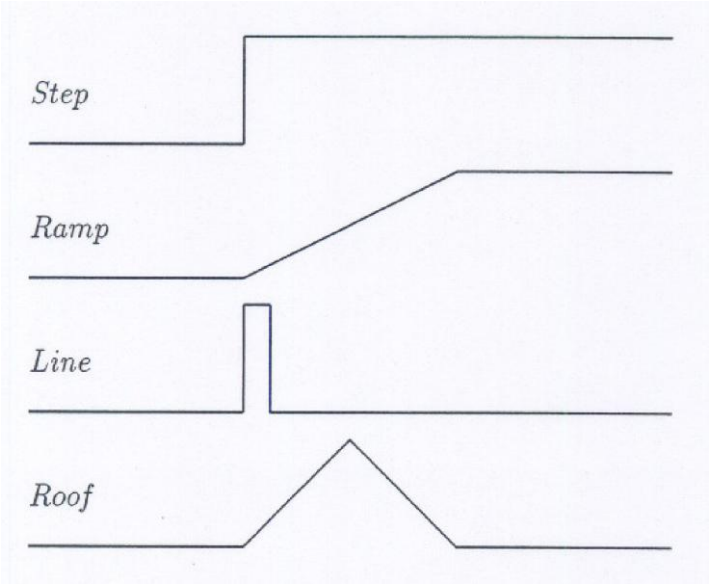
- **Ακμή Γραμμής (Line Edge):**
  - Στενές κορυφογραμμές ή λωρίδες όπου η φωτεινότητα ανεβαίνει και κατεβαίνει απότομα.
  - Παράδειγμα: μια λεπτή άσπρη γραμμή πάνω σε σκοτεινό φόντο.

Στην πράξη, οι ιδανικές ακμές τύπου βήματος και γραμμής εμφανίζονται σπάνια λόγω της επίδρασης φυσικών φαινομένων όπως η σκέδαση του φωτός, ο θόρυβος αισθητήρων και της εξομάλυνση που εισάγεται από τις περισσότερες συσκευές ανίχνευσης. Συνεπώς, οι ακμές αυτές εμφανίζονται πιο συχνά υπό τις μορφές ράμπας και οροφής, αντίστοιχα:

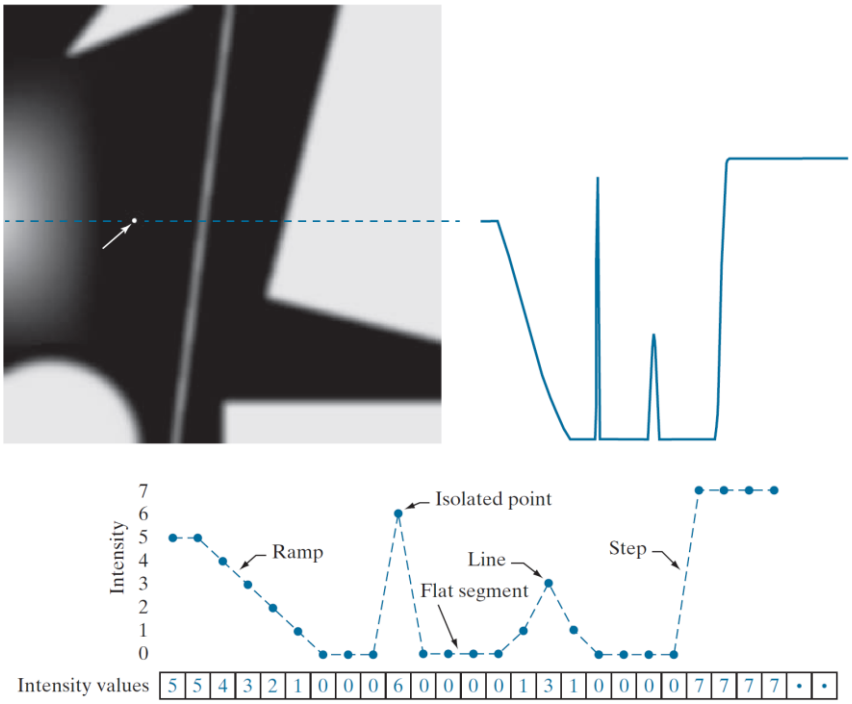
- **Ακμή Ράμπας (Ramp Edge):**
  - Ομαλή και σταδιακή μεταβολή της φωτεινότητας από τη μία τιμή στην άλλη.
  - Παράδειγμα: η σκιά που δημιουργείται από μια κυρτή επιφάνεια, όπου το φως σταδιακά μειώνεται από φωτεινό σε σκοτεινό.

- **Ακμή Οροφής (Roof Edge):**
  - Η φωτεινότητα αυξάνεται ομαλά, φτάνει σε μια μέγιστη ένταση και στη συνέχεια μειώνεται εξίσου ομαλά.
  - Παράδειγμα: στενές κορυφογραμμές ή φωτεινές λωρίδες, όπως μια φωτεινή γραμμή επάνω σε πιο σκοτεινό φόντο, όπου η μετάβαση γίνεται σταδιακά και όχι απότομα.

Τέλος, οι παραπάνω τύποι ακμών, οι οποίοι φαίνονται στο σχήμα 3.1, μπορούν να συνυπάρχουν ή να αλληλοεπιδρούν μεταξύ τους και συχνά επηρεάζονται από την παρουσία θορύβου, γεγονός που καθιστά την ακριβή ανίχνευσή τους σε πραγματικές συνθήκες ιδιαίτερα απαιτητική όπως αυτή του σχήματος 3.2.



Σχήμα 3.1: Απεικόνιση των βασικών τύπων ακμών: βήματος, ράμπας, γραμμής και οροφής [4]



Σχήμα 3.2: Ένα παράδειγμα εικόνας μαζί με τις τιμές φωτεινότητας (έντασης) κατά μήκος της μπλε διακεκομμένης γραμμής, όπου απεικονίζονται οι διάφοροι τύποι ακμών [5]

3.1.2.2 Ανιχνευτής Ακμών

Εξ ορισμού, ο ανιχνευτής ακμών είναι ένας αλγόριθμος που εξάγει από μια εικόνα ένα σύνολο ακμών, είτε ως μεμονωμένα σημεία είτε ως συνεχόμενα τμήματα [4]. Οι ακμές αποτελούν βασικά χαρακτηριστικά μιας εικόνας, καθώς συνήθως αντιστοιχούν σε σημαντικές δομές όπως όρια αντικειμένων, μεταβολές υφής ή διαφοροποιήσεις φωτεινότητας. Η ακριβής και αποδοτική ανίχνευσή τους είναι κρίσιμη σε ποικίλες εφαρμογές της ψηφιακής επεξεργασίας εικόνας, όπως στην ιατρική απεικόνιση, στη ρομποτική όραση και στην αναγνώριση αντικειμένων.

Οι κύριες προκλήσεις γίνονται εμφανείς μέσα από την έννοια ενός «ιδανικού ανιχνευτή ακμών» [1]. Ένας τέτοιος ανιχνευτής θα πρέπει να επιτυγχάνει ταυτόχρονα υψηλή ακρίβεια και υψηλή απόδοση. Όσον αφορά την ακρίβεια, αυτό συνεπάγεται ότι:

1. Ανιχνεύει τα πραγματικά περιγράμματα με τη μέγιστη δυνατή ακρίβεια, ελαχιστοποιώντας ή εξαλείφοντας τις ψευδείς ακμές.
2. Τοποθετεί κάθε σημείο ακμής στη σωστή θέση, χωρίς σφάλματα εντοπισμού.
3. Παραμένει ανθεκτικός στον θόρυβο, ανεξάρτητα από την προέλευσή του, είτε προκύπτει από την ίδια την εικόνα είτε από φίλτρα και άλλες διαδικασίες επεξεργασίας.

Αντίστοιχα, σε όρους απόδοσης, ένας ανιχνευτής ακμών πρέπει να χαρακτηρίζεται από χαμηλές υπολογιστικές απαιτήσεις σε επίπεδο υλικού (hardware) και λογισμικού (software), ώστε να είναι εφαρμόσιμος σε πραγματικό χρόνο και προσβάσιμος σε διαφορετικά περιβάλλοντα και εφαρμογές. Το αποτέλεσμα ενός ανιχνευτή ακμών μπορεί να ταξινομηθεί σε τέσσερις κατηγορίες [4]:

- **Αληθώς θετικές (true positives):** Πρόκειται για τις περιπτώσεις όπου ο ανιχνευτής εντοπίζει σωστά τις πραγματικές ακμές της εικόνας.
- **Ψευδώς θετικές (false positives):** Περιπτώσεις όπου ο ανιχνευτής αναγνωρίζει ακμές που στην πραγματικότητα δεν υπάρχουν, δημιουργώντας λανθασμένα αποτελέσματα..
- **Ψευδώς αρνητικές (false negatives):** Όταν ο ανιχνευτής αποτυγχάνει να εντοπίσει υπαρκτές ακμές της εικόνας, με αποτέλεσμα την απώλεια σημαντικής πληροφορίας.
- **Αληθώς αρνητικές (true negatives):** Οι περιοχές της εικόνας όπου δεν υπάρχουν ακμές και ο ανιχνευτής σωστά δεν εντοπίζει καμία ακμή.

Η αξιολόγηση ενός ανιχνευτή βασίζεται στη σχέση ανάμεσα σε αυτές τις κατηγορίες και στο ποσοστό επιτυχίας που επιτυγχάνει στον διαχωρισμό πραγματικών από ψευδείς ακμές. Ο τελικός στόχος είναι η μεγιστοποίηση της ακρίβειας και η βελτίωση της αξιοπιστίας του συστήματος σε πρακτικές εφαρμογές.

3.1.2.3 Μέθοδοι Αναγνώρισης Ακμών

Η αναγνώριση ακμών αποτελεί θεμελιώδη διαδικασία στην επεξεργασία εικόνας, καθώς επιτρέπει τον εντοπισμό σημαντικών δομών, όπως τα όρια μεταξύ αντικειμένων ή περιοχών διαφορετικών χαρακτηριστικών. Οι βασικές μέθοδοι ανίχνευσης ακμών στηρίζονται στην ανάλυση της έντασης της εικόνας. Η κεντρική τους αρχή είναι ότι οι ακμές εντοπίζονται σε περιοχές όπου η φωτεινότητα μεταβάλλεται απότομα, υποδηλώνοντας τη μετάβαση από μια περιοχή σε μια άλλη. Τα βασικά στάδια ανίχνευσης ακμών περιλαμβάνουν [4,5]:

- **Υπολογισμό της κλίσης της εικόνας** μέσω παραγώγων πρώτης τάξης (π.χ. τελεστές Roberts, Prewitt, Sobel) ή δεύτερης τάξης (π.χ. Laplacian, Laplacian of Gaussian).

- **Εφαρμογή μάσκας συνέλιξης (convolution mask)**, η οποία λειτουργεί ως φίλτρο και εκτιμά τις τοπικές μεταβολές της έντασης σε οριζόντιες, κάθετες ή και διαγώνιες κατευθύνσεις.
- **Χρήση κατωφλίων (thresholding)** ή άλλων κριτηρίων, με στόχο τον διαχωρισμό των "ισχυρών" ακμών από τον θόρυβο ή τις ασθενείς ακμές που δεν αντιστοιχούν σε πραγματικές δομές της εικόνα.

Για την εφαρμογή των περισσοτέρων τελεστών ακμών, η έγχρωμη εικόνα μετατρέπεται αρχικά σε αποχρώσεις του γκρι, διατηρώντας μόνο την πληροφορία φωτεινότητας. Η διαδικασία αυτή μειώνει τη διάσταση του προβλήματος και επιτρέπει τη χρήση αριθμητικών τελεστών για τον υπολογισμό των μεταβολών.

Παρά το γεγονός ότι αυτές οι μέθοδοι είναι απλές, εύκολα υλοποιήσιμες και υπολογιστικά αποδοτικές, παρουσιάζουν ορισμένους περιορισμούς: είναι ευαίσθητες στον θόρυβο, απαιτούν προσεκτική επιλογή κατωφλίων και σε πολλές περιπτώσεις εντοπίζουν πολλαπλά σημεία ακμής αντί για μία καθαρή γραμμή. Ωστόσο, εξακολουθούν να χρησιμοποιούνται, ιδίως σε εφαρμογές όπου απαιτείται χαμηλή υπολογιστική πολυπλοκότητα.

3.1.2.4 Κλίση της Εικόνας

Μια εικόνα μπορεί να θεωρηθεί ως ένας πίνακας διακριτών δειγμάτων μιας συνεχούς συνάρτησης $f(x,y)$, η οποία περιγράφει τις τιμές έντασης φωτεινότητας σε κάθε σημείο της εικόνας [4]. Μεγάλες αλλαγές στις τιμές έντασης μιας εικόνας μπορούν να ανιχνευτούν μέσω μίας διακριτής προσέγγιση της κλίσης (gradient). Για παράδειγμα , στην περίπτωση μιας ακμής βήματος (step edge), η ακμή αντιστοιχεί σε μια τοπική κορυφή στην πρώτη παράγωγο. Ο εντοπισμός των ακμών μπορεί να πραγματοποιηθεί είτε μέσω παραγώγων πρώτης τάξης είτε μέσω παραγώγων δεύτερης τάξης [5]. Η κλίση σε ένα σημείο (x,y), της εικόνας f ορίζεται ως το διάνυσμα [4,5]:

$$\nabla f \;=\; grad[f(x,y)] = \begin{bmatrix} g_x(x,y) \\ g_y(x,y) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f(x,y)}{\partial x} \\ \dfrac{\partial f(x,y)}{\partial y} \end{bmatrix} \qquad (3.1.1)$$

Το μέτρο - μήκος (magnitude) του διανύσματος, $M(x,y)$, του διανύσματος της κλίσης στο σημείο (x,y), ορίζεται ως:

$$M(x,y) = \;\| \, grad[f(x,y)] \, \| = \sqrt{\; g_x^2(x,y) + \; g_y^2(x,y)} \qquad (3.1.2)$$

Μια υπολογιστικά ελαφρύτερη προσέγγιση, η οποία διατηρεί τις σχετικές μεταβολές στα επίπεδα έντασης, είναι η εξής:

$$M(x,y) \approx \; |g_x(x,y)| + \; \left|g_y(x,y)\right| \qquad (3.1.3)$$

Προκύπτουν δύο σημαντικές ιδιότητες από την κλίση της εικόνας:

- Το διάνυσμα ∇f δείχνει την διεύθυνση του μεγαλύτερου ρυθμού μεταβολής της f στο σημείο (x,y).
- Το μέτρο $M(x,y)$ είναι η τιμή του ρυθμού μεταβολής κατά την διεύθυνση του διανύσματος κλίσης και περιγράφει πόσο απότομη είναι η τοπική αλλαγή της ένταση στο σημείο (x,y).

Όταν για κάθε εικονοστοιχείο (pixel) μπορούν να υπολογιστούν οι συναρτήσεις $g_x(x,y)$, $g_y(x,y)$ και $M(x,y)$, τότε οι παραγόμενες εικόνες έχουν το ίδιο μέγεθος με την αρχική εικόνα f.

Η διεύθυνση του διανύσματος κλίσης σε ένα σημείο (x, y) προσδιορίζεται από την γωνία:

$$a(x,y) = \tan^{-1}\left(\frac{g_y(x,y)}{g_x(x,y)}\right) \quad (3.1.4)$$

Η γωνία a υπολογίζεται αριστερόστροφα ως προς τον άξονα x. Η διεύθυνση μιας ακμής στο σημείο (x, y) είναι κάθετη προς τη διεύθυνση του διανύσματος κλίσης σε αυτό το σημείο. Επομένως, η διεύθυνση της ακμής στο σημείο (x, y) προσδιορίζεται από την γωνία $a(x, y) - 90^o$.

Για τον υπολογισμό της κλίσης ∇f μιας εικόνας υπολογίζουμε τις μερικές παραγωγούς $\partial f/\partial x$ και $\partial f/\partial y$ για κάθε εικονοστοιχείο της εικόνας [4,5]. Οι πρώτες παράγωγοι σε μια γειτονιά γύρω από το σημείο (x, y) μπορούν να προσεγγιστούν με διαφορές, σύμφωνα με τις σχέσεις:

$$g_x(x,y) = \frac{\partial f(x,y)}{\partial x} = f(x+1,y) - f(x,y) \quad (3.1.5)$$

$$g_y(x,y) = \frac{\partial f(x,y)}{\partial y} = f(x,y+1) - f(x,y) \quad (3.1.6)$$

Αυτές οι δύο εξισώσεις υπολογίζονται για κάθε εικονοστοιχείο της εικόνας εφαρμόζονται μονοδιάστατες μάσκες (1-D kernels):

$g_x =$ | -1 | 1 | $g_y =$ | -1 |
|---|---|---|---|
| | | | 1 |

Στην πράξη χρησιμοποιούνται μάσκες που είναι συμμετρικές ως προς το κέντρο τους, η μικρότερη εκ των οποίων έχει διαστάσεις $3\ x\ 3$. Με αυτόν τον τρόπο, η τιμή της κλίσης προκύπτει με αναφορά στο κεντρικό εικονοστοιχείο της γειτονιάς. Έστω η παρακάτω μάσκα $3\ x\ 3$ με τις τιμές ε να αντιστοιχούν σε τιμές έντασης:

ε_1	ε_2	ε_3
ε_4	ε_5	ε_6
ε_7	ε_8	ε_9

Οι ψηφιακές προσεγγίσεις των μερικών παραγώγων για το κεντρικό σημείο, δίδονται από τις παρακάτω σχέσεις:

$$g_x = \frac{\partial f}{\partial x} = (\varepsilon_7 + \varepsilon_8 + \varepsilon_9) - (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) \quad (3.1.7)$$

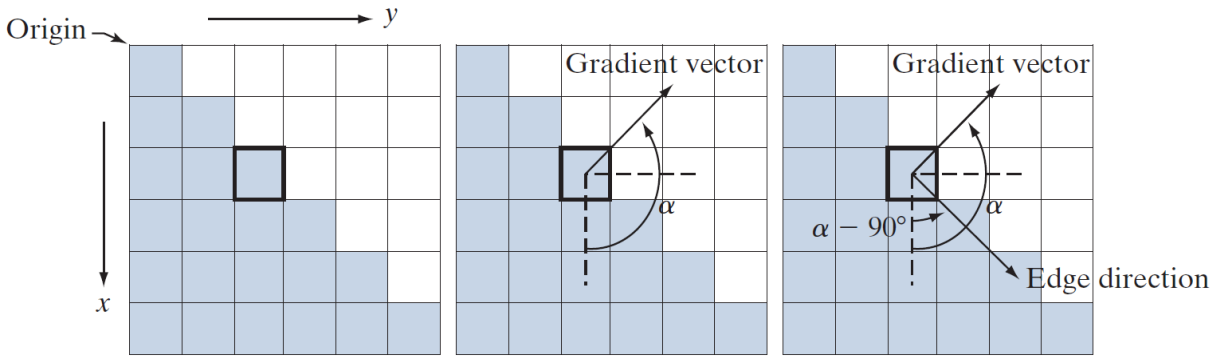
$$g_y = \frac{\partial f}{\partial y} = (\varepsilon_3 + \varepsilon_6 + \varepsilon_9) - (\varepsilon_1 + \varepsilon_4 + \varepsilon_7) \quad (3.1.8)$$

Άρα η διαφορά ανάμεσα στην τρίτη και στην πρώτη γραμμή της μάσκας $3\,x\,3$, αποτελεί μια προσέγγιση της τιμής της μερικής παραγώγου $\partial f/\partial x$ και η διαφορά ανάμεσα στην τρίτη και στην πρώτη στήλη, προσεγγίζει την τιμής της μερικής παραγώγου $\partial f/\partial y$.

Χαρακτηριστικό παράδειγμα [5] αποτελεί το σχήμα 3.3, η εικόνα f είναι σε κλίμακα του γκρι, όπου τα γκρι εικονοστοιχεία αντιστοιχούν σε μηδενικές τιμές και τα άσπρα σε τιμές ίσες με την μονάδα. Για τον υπολογισμό των παραγώγων, χρησιμοποιείται μια τοπική γειτονιά διαστάσεων $3\,x\,3$, με κέντρο το εικονοστοιχείο που επισημαίνεται με μαύρο πλαίσιο.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

με μήκος $M(x,y) = 2\sqrt{2}$ και γωνία $a(x,y) = -45^o$, που ισοδυναμεί με 135^o όταν μετρηθεί ως θετική (αριστερόστροφη) γωνία ως προς τον άξονα x, σχήμα 3.3 (β). Όπως αναφέρθηκε προηγουμένως, η διεύθυνση της ακμής είναι κάθετη στη διεύθυνση του διανύσματος κλίσης. Επομένως, προσδιορίζεται αριστερόστροφά από τον θετικό άξονα x με βάση τη γωνία $a - 90^o = 135^o - 90^o = 45^o$, σχήμα 3.3 (γ).



(α) (β) (γ)

Σχήμα 3.3: Υπολογισμός του μεγέθους και της διεύθυνσης της ακμής μέσω της εφαρμογής του διανύσματος κλίσης (gradient vector) στην αρχική εικόνα (α), εστιάζοντας στο εικονοστοιχείο που επισημαίνεται με μαύρο πλαίσιο [5]

3.1.2.5 Τελεστές Κλίσης - Πρώτης Παραγώγου

Οι πιο διαδεδομένοι τελεστές πρώτης παραγώγου είναι οι Roberts, Prewitt και Sobel, οι οποίοι εφαρμόζονται για τον υπολογισμό της κλίσης της έντασης της εικόνας, με στόχο την ανίχνευση ακμών. Οι τελεστές αυτοί διαφέρουν ως προς τη μορφή των μασκών συνέλιξης, την ευαισθησία τους στον θόρυβο και την ικανότητά τους να εντοπίζουν διαφορετικού τύπου ακμές [4,5]. Η εφαρμογή των τελεστών πραγματοποιείται μέσω της διαδικασίας της συνέλιξης (convolution), κατά την οποία ένα παράθυρο (μάσκα) μετακινείται διαδοχικά πάνω από κάθε θέση της εικόνας. Σε κάθε σημείο, υπολογίζεται μια νέα τιμή για το αντίστοιχο εικονοστοιχείο, βασισμένη στα γειτονικά του σημεία, σύμφωνα με τους συντελεστές της μάσκας.

Οι τελεστές Roberts αποτελούν από τις πρώτες προσπάθειες χρήσης δισδιάστατων μασκών (2-D kernels) που επικεντρώνονται στην διαγώνια διεύθυνση. Έστω η μάσκα $3\,x\,3$ με τις τιμές έντασης ε,

που ορίσαμε στο προηγούμενο κεφάλαιο. Οι τελεστές Roberts υπολογίζοντας βρίσκοντας τις διαγώνιες διαφορές, για τον υπολογισμό της κλίσης στο σημείο ε_5:

$$g_x = \frac{\partial f}{\partial x} = \varepsilon_9 - \varepsilon_5 \quad (4.1)$$

$$g_y = \frac{\partial f}{\partial y} = \varepsilon_8 - \varepsilon_6 \quad (4.2)$$

Για την εφαρμογής τους φιλτράρουμε μια εικόνα με τις παρακάτω μάσκες:

$g_x =$

-1	0
0	1

$g_y =$

0	-1
1	0

Παρά την απλότητα τους, οι μάσκες διαστάσεων $2\ x\ 2$ δεν χρησιμοποιούνται ευρέως στην ανίχνευση ακμών, καθώς δεν διαθέτουν σαφές κεντρικό εικονοστοιχείο στο οποίο να αντιστοιχεί η τιμή της παραγώγου. Αυτό έχει ως αποτέλεσμα να καθίσταται δυσκολότερος ο ακριβής προσδιορισμός της θέσης της ακμής, σε σύγκριση με μάσκες $3\ x\ 3$ ή μεγαλύτερες, οι οποίες παρέχουν καλύτερη ακρίβεια χωρικού εντοπισμού. Επιπλέον, η μικρή τους διάσταση τις καθιστά ιδιαίτερα ευαίσθητες στον θόρυβο.

Για τον λόγο αυτό, στην πράξη προτιμώνται μάσκες που είναι συμμετρικές ως προς το κέντρο τους, την ελάχιστη διάστασή τους να είναι $3\ x\ 3$. Οι μάσκες αυτού του τύπου επιτρέπουν τον υπολογισμό της παραγώγου με αναφορά στο κεντρικό εικονοστοιχείο. Μια απλοποιημένη αλλά αποτελεσματική υλοποίηση τέτοιων τελεστών είναι οι τελεστές Prewitt, οι οποίοι χρησιμοποιούν μη σταθμισμένες διαφορές για την προσέγγιση των μερικών παραγώγων κατά τις διευθύνσεις x και y. Οι μερικές παράγωγοι περιγράφονται από τις εξισώσεις (4.7) και (4.8), ενώ οι αντίστοιχες μάσκες έχουν την ακόλουθη μορφή:

$g_x =$

-1	-1	-1
0	0	0
1	1	1

$g_y =$

-1	0	1
-1	0	1
-1	0	1

Οι μάσκες αυτές εφαρμόζονται μέσω συνέλιξης πάνω στην εικόνα, επιτρέποντας την εκτίμηση της τοπικής κλίσης σε κάθε εικονοστοιχείο.

Μια μικρή παραλλαγή των εξισώσεων (4.7) και (4.8), στις οποίες δίνεται έμφαση στην κεντρική γραμμή της μάσκας, οδηγεί στους τελεστές Sobel. Σε αυτή την περίπτωση, οι μάσκες δίνουν μεγαλύτερη βαρύτητα στους κεντρικούς συντελεστές, πολλαπλασιάζοντάς τους με βάρος ίσο με 2:

$$g_x = \frac{\partial f}{\partial x} = (\varepsilon_7 + 2\varepsilon_8 + \varepsilon_9) - (\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3) \quad (4.3)$$

$$g_y = \frac{\partial f}{\partial y} = (\varepsilon_3 + 2\varepsilon_6 + \varepsilon_9) - (\varepsilon_1 + 2\varepsilon_4 + \varepsilon_7) \quad (4.4)$$

Αυτός ο σταθμισμένος υπολογισμός της κλίσης καθιστά τους τελεστές Sobel πιο ευαίσθητους στις αλλαγές έντασης, ενώ παράλληλα βελτιώνει την ανθεκτικότητά τους στον θόρυβο συγκριτικά με τους μη σταθμισμένους τελεστές, όπως οι Prewitt. Οι μάσκες Sobel ορίζονται ως εξής:

$g_x =$

-1	-2	-1
0	0	0
1	2	1

$g_y =$

-1	0	1
-2	0	2
-1	0	1

Επιπλέον, οι τελεστές Sobel (καθώς και οι τελεστές Prewitt) μπορούν να προσαρμοστούν ώστε να ανιχνεύουν και διαγώνιες ακμές, πέραν των οριζόντιων και κατακόρυφων. Αυτό επιτυγχάνεται με την κατασκευή κατάλληλων μασκών που είναι ευαίσθητες σε συγκεκριμένες διαγώνιες κατευθύνσεις, όπως οι 45° και 135°:

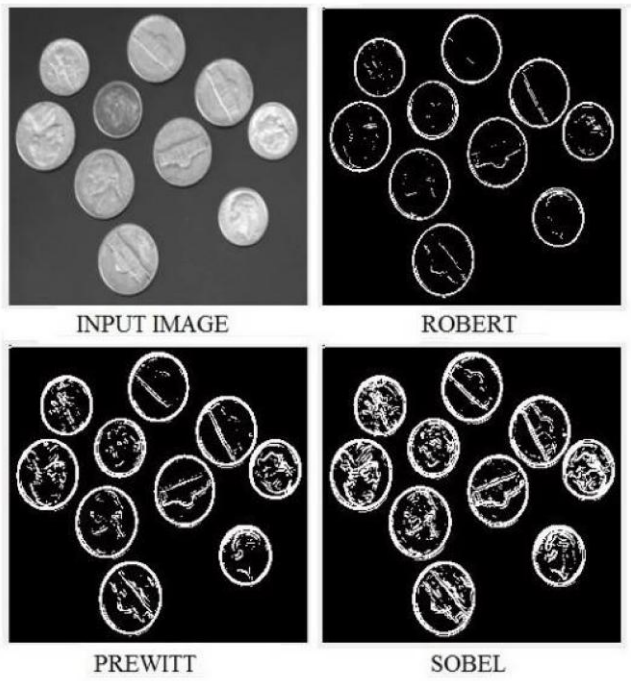
$g_x =$

0	1	2
-1	0	1
-2	-1	0

$g_y =$

-2	-1	0
-1	0	1
0	1	2

Με αυτό τον τρόπο ενισχύεται η δυνατότητα του ανιχνευτή να ανταποκρίνεται σε πιο σύνθετες γεωμετρικές δομές μέσα στην εικόνα. Να σημειωθεί ότι και στις τρεις περιπτώσεις τελεστών, Roberts, Prewitt και Sobel, το άθροισμα των συντελεστών τους ισούται με μηδέν. Επομένως, σε περιοχές όπου η εικόνα έχει σταθερή τιμή έντασης (δηλαδή δεν υπάρχουν μεταβολές φωτεινότητας ή υφής), η απόκριση των τελεστών είναι μηδενική.



Σχήμα 3.4: Σύγκριση απόκρισης των τελεστών Roberts, Prewitt και Sobel στην ανίχνευση ακμών [6]

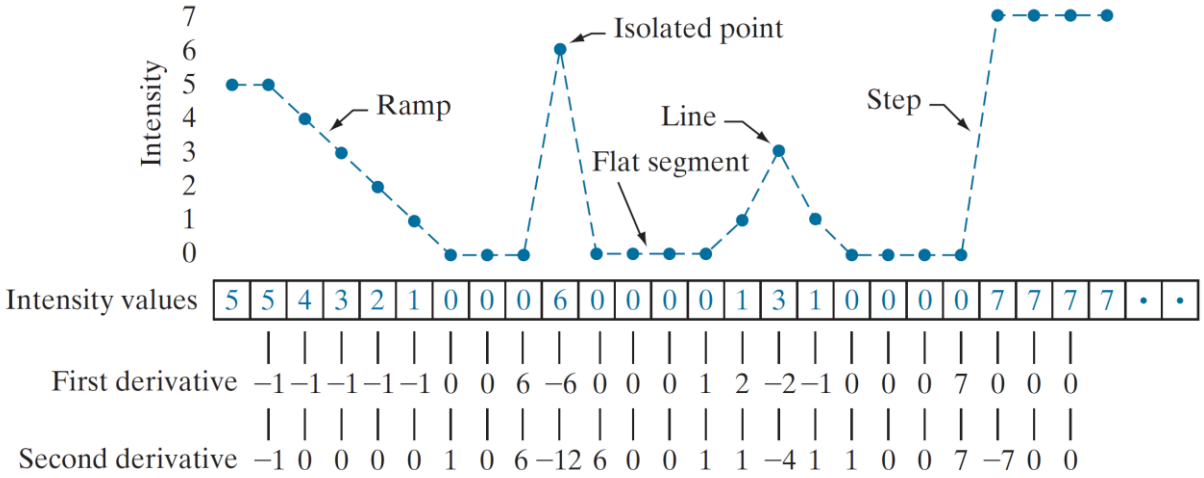
3.1.2.6 Τελεστές Κλίσης - Δεύτερης Παραγώγου

Η πρώτη παράγωγος χρησιμοποιείται για τον εντοπισμό περιοχών όπου η ένταση αλλάζει γρήγορα, δηλαδή των ακμών, παρέχοντας πληροφορία για τη διεύθυνση της μεταβολής. Αντίθετα, η δεύτερη παράγωγος μετρά τον ρυθμό μεταβολής της κλίσης, επιτρέποντας τον ακριβή εντοπισμό κορυφών, ορίων ακμών και λεπτομερειών. Παράλληλα, παρουσιάζει ισχυρότερη απόκριση σε λεπτές γραμμές, απομονωμένα σημεία και θόρυβο, ενώ παράγει διπλή απόκριση (double-edge response) σε μεταβάσεις βήματος ή ράμπας και το πρόσημο της μπορεί να χρησιμοποιηθεί για να καθοριστεί η κατεύθυνση της μετάβασης από φωτεινό σε σκοτεινό ή αντίστροφα [5]. Οι δεύτερες μερικές παράγωγοι της κλίσης σε ένα σημείο (x, y), της εικόνας f ορίζονται ως [4,5]:

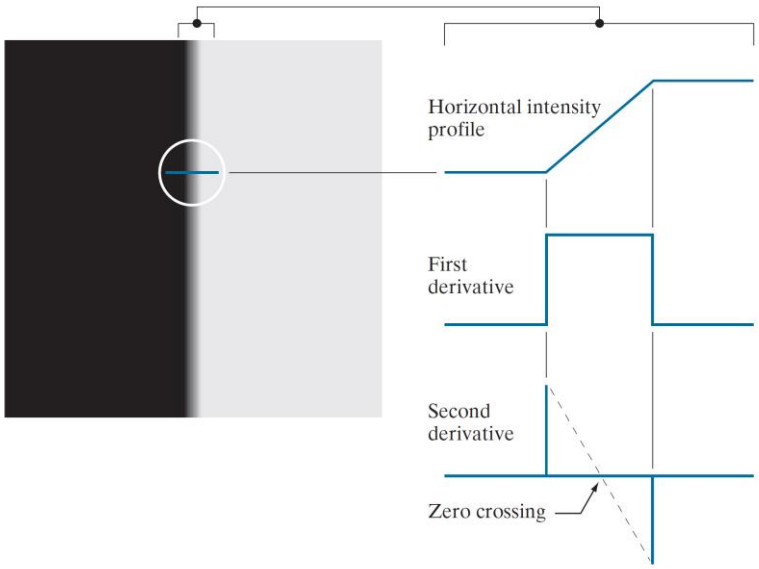
$$\frac{\partial^2 f(x,y)}{\partial x^2} = f(x-1,y) - 2f(x,y) + f(x+1,y) \quad (3.1.9)$$

$$\frac{\partial^2 f(x,y)}{\partial y^2} = f(x,y+1) - 2f(x,y) + f(x,y-1) \quad (3.1.10)$$

Για παράδειγμα, έστω μια ακμή βήματος (step edge) με τιμές έντασης από 0 σε 7. Η εφαρμογή της δεύτερης παραγώγου ως προς τον άξονα x τη μετασχηματίζει ε μια ακολουθία τιμών που διασχίζουν το μηδέν (zero crossing), παρουσιάζοντας θετική κορυφή στο +7 και αρνητική στο -7, όπως φαίνεται στο σχήμα 3.5. Το σημείο της μετάβασης από θετικές σε αρνητικές τιμές υποδεικνύει τη θέση της ακμής, η οποία μπορεί να προσδιοριστεί με μεγαλύτερη ακρίβεια ενδιάμεσα στα δύο εικονοστοιχεία χρησιμοποιώντας γραμμική παρεμβολή (linear interpolation). Στην πράξη, επιλέγεται το εικονοστοιχείο αριστερά ή δεξιά της ακμής, αρκεί να γίνεται συνεπής επιλογή σε ολόκληρη την εικόνα. Παρόμοια, σε μια ακμή ράμπας (ramp edge) όπου η ένταση μειώνεται σταδιακά (σχήμα 3.6), η δεύτερη παράγωγος παράγει επίσης μια διπλή κορυφή: μια αρνητική και μια θετική, στην αρχή και το τέλος της μετάβασης αντίστοιχα, υποδεικνύοντας με μεγαλύτερη ακρίβεια τα όρια της ακμής.



Σχήμα 3.5: Απεικόνιση πρώτης και δεύτερης παραγώγου για διαφορετικούς τύπους ακμών [5]



Σχήμα 3.6: Απεικόνιση της πρώτης και της δεύτερης παραγώγου για μια ακμή τύπου ράμπας [5]

Η συνάρτηση του Laplacian, ισοδύναμη με το άθροισμα των δεύτερων μερικών παραγώγων, ενισχύει τον εντοπισμό περιοχών με απότομες μεταβολές της έντασης, καθιστώντας την ιδιαίτερα χρήσιμη για τον ακριβή εντοπισμό ακμών και λεπτομερειών. Το άθροισμα των δεύτερων μερικών παραγώγων εκφράζεται μαθηματικά ως:

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \qquad (3.1.11)$$

Συνδυάζοντας τις εξισώσεις (4.13) και (4.14) με την (4.15) παίρνουμε:

$$\nabla^2 f(x,y) = f(x-1,y) + f(x+1,y) + f(x,y-1) + f(x,y+1) - 4f(x,y) \quad (3.1.12)$$

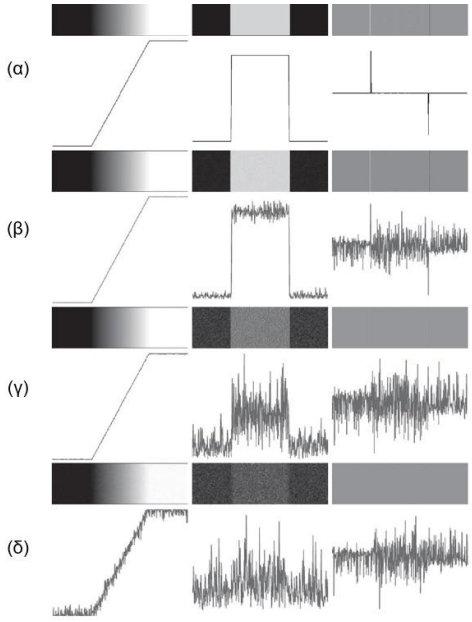
Η αντίστοιχη $3\ x\ 3$ Laplacian μάσκα είναι η:

0	1	0
1	-4	1
0	1	0

Αν θέλουμε να δώσουμε μεγαλύτερη βαρύτητα στο κεντρικό εικονοστοιχείο, τότε χρησιμοποιηθούμε μια προσέγγιση της Laplacian μάσκας:

1	4	1
4	-20	4
1	4	1

Οι τελεστές Laplacian και οι τελεστές δεύτερης κατεύθυνσης δεν χρησιμοποιούνται συχνά στην πράξη καθώς είναι περισσότερο ευαίσθητοι στον θόρυβο σε σχέση με έναν τελεστή πρώτης παραγώγου. Ένα παράδειγμα για το πως επηρεάζει ο θόρυβός την ανίχνευση μιας ακμής ράμπας απεικονίζεται στο σχήμα 3.7. Για την αντιμετώπιση αυτού του προβλήματος, είναι αναγκαία η εφαρμογή τεχνικών προ-φιλτραρίσματος της εικόνας, όπως η χρήση φίλτρου Gaussian, με στόχο τη μείωση της επίδρασης του θορύβου και τη βελτίωση της αξιοπιστίας στον εντοπισμό ακμών.



Σχήμα 3.7**:** Επίδραση του θορύβου στην ανίχνευση μιας ακμής ράμπας. Στην πρώτη στήλη απεικονίζεται η αρχική εικόνα με θόρυβο Gaussian μηδενικής μέσης τιμής και τυπικών αποκλίσεων 0.0, 0.1, 1.0 και 10.0 αντίστοιχα για τις εικόνες (α), (β), (γ) και (δ). Στη

δεύτερη στήλη φαίνονται οι εικόνες της πρώτης παραγώγου και τα προφίλ έντασης, ενώ στην τρίτη στήλη παρουσιάζονται οι εικόνες της δεύτερης παραγώγου και τα αντίστοιχα προφίλ έντασης [5]

Ο Marr και Hildreth [4,5] πρότειναν μια μέθοδο συνδυασμού της εξομάλυνσης της εικόνας με Gaussian φίλτρο και της ανίχνευσης ακμών μέσω δεύτερης παραγώγου, γνωστή ως Laplacian of Gaussian (LoG). Η βασική ιδέα πίσω από τον LoG είναι να μειώνεται η επίδραση του θορύβου πριν τον εντοπισμό ακμών, ενώ ταυτόχρονα να ενισχύονται οι περιοχές όπου η ένταση της εικόνας αλλάζει απότομα. Η μαθηματική έκφραση του LoG δίνεται από:

$$\nabla^2 G(x,y) = \left[\frac{x^2+y^2-2\sigma^2}{\sigma^4}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (3.1.13)$$

όπου:

- ∇^2 αναφέρετε στον τελεστή Laplace, δηλαδή το άθροισμα των δεύτερων μερικών παραγώγων
- Η συνάρτηση Gauss με τυπική απόκλιση *σ* δίδεται από την εξίσωση:

$$G(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (3.1.14)$$

Η εφαρμογή του αλγορίθμου Marr - Hildreth σε μια εικόνα $f(x,y)$ πραγματοποιείται μέσω της συνέλιξης του φίλτρου LoG με την εικόνα:

$$g(x,y) = [\nabla^2 G(x,y)] * f(x,y) \qquad (3.1.15)$$

Ο προσδιορισμός των θέσεων των ακμών στην εικόνα $f(x,y)$, επιτυγχάνεται μέσω της ανίχνευσης των σημείων όπου γίνεται η διέλευση από το μηδέν της συνάρτησης $g(x,y)$, τα οποία αντιστοιχούν σε απότομες μεταβολές της έντασης και καθορίζουν με ακρίβεια την θέση των ακμών. Εναλλακτικά, χωρίς να επηρεαστεί το τελικό αποτέλεσμα, η εξίσωση (4.19) μπορεί να γραφεί και ως:

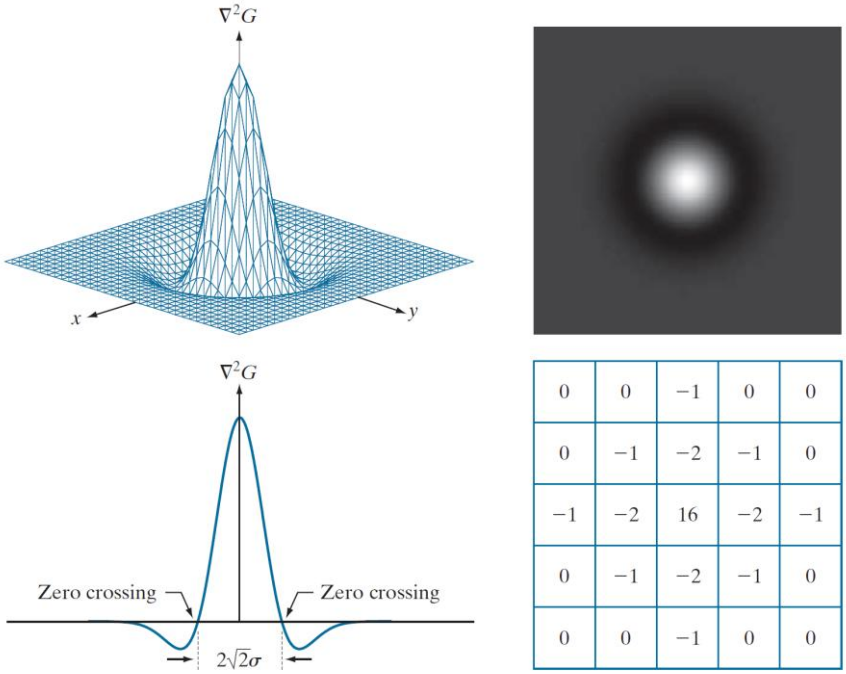
$$g(x,y) = \nabla^2 [G(x,y) * f(x,y)] \qquad (3.1.16)$$

υποδηλώνοντας πως ότι πρώτα εξομαλύνουμε την εικόνα με το φίλτρο Gaussian για να μειώσουμε τον θόρυβο και στη συνέχεια εφαρμόζουμε τον τελεστή Laplace. Συνοψίζοντας, ο αλγόριθμος Marr-Hildreth για την ανίχνευση ακμών εκτελείται σε τρία βασικά βήματα [5]:

1. **Εξομάλυνση της εικόνας με φίλτο Gaussian:**
   Η αρχική εικόνα φιλτράρεται με Gaussian για μείωση του θορύβου και αποφυγή ψευδών ακμών.

2. **Εφαρμογή του Laplacian**
   Στην εξομαλυμένη εικόνα εφαρμόζεται ο τελεστής Laplace, ενισχύοντας τις περιοχές με απότομες μεταβολές της έντασης.

3. **Εντοπισμός σημείων**
   Οι θέσεις των ακμών προσδιορίζονται από τα σημεία διέλευσης από το μηδέν, δηλαδή τα σημεία όπου η συνάρτηση Laplacian της εικόνας που προέκυψε από το βήμα 2 αλλάζει πρόσημο.



0	0	−1	0	0
0	−1	−2	−1	0
−1	−2	16	−2	−1
0	−1	−2	−1	0
0	0	−1	0	0

Σχήμα 3.8: Πρώτη γραμμή: Τρισδιάστατο γράφημα (αριστερά) και απεικόνισή του ως εικόνα (δεξιά) για το αρνητικό της συνάρτησης LoG. Δεύτερη γραμμή: : Διατομή του τρισδιάστατου γραφήματος με επισήμανση των σημείων διέλευσης από το μηδέν (zero-crossing, αριστερά) και αντίστοιχη μάσκα διαστάσεων $\mathbf{5}\ \boldsymbol{x}\ \mathbf{5}$ (δεξιά). Το αρνητικό αυτής της μάσκας χρησιμοποιείται στην πράξη [5]

3.1.2.7 Ο ανιχνευτής Ακμών Canny

Ο Ανιχνευτής ακμών Canny [4,5] είναι ένας πιο περίπλοκος αλγόριθμος που υπερτερεί από σε απόδοση από τις προηγούμενους αλγορίθμους. Η προσέγγιση του Canny βασίζεται σε τρεις βασικούς στόχους:

1. **Καλός Εντοπισμός:** Όλες οι πραγματικές ακμές πρέπει να ανιχνεύονται, χωρίς ψευδείς ανιχνεύσεις.

2. **Καλή Τοποθέτηση:** Τα σημεία ακμών που εντοπίζονται πρέπει να βρίσκονται όσο το δυνατόν πιο κοντά στις πραγματικές ακμές.

3. **Απόκριση Μοναδικού Σημείου:** Ο ανιχνευτής πρέπει να επιστρέφει μόνο ένα σημείο για κάθε πραγματική ακμή. Αυτό σημαίνει ότι το πλήθος των τοπικών μεγίστων γύρω από την πραγματική ακμή πρέπει να είναι ελάχιστος, , ώστε να αποφεύγεται η ανίχνευση πολλαπλών σημείων ακμής εκεί όπου υπάρχει μόνο ένα πραγματικό σημείο ακμής.

Έστω η εικόνας εισόδου $f(x,y)$, το πρώτο βήμα του αλγορίθμου Canny είναι να εξομαλύνει την εικόνα με ένα φίλτρο Gauss $G(x,y)$, με στόχο τη μείωση του θορύβου:

$$f_s(x,y) = G(x,y) * f(x,y) \quad \textit{και} \quad G(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.1.17}$$

Όπου το $*$ δηλώνει πράξη συνέλιξη (convolution) και η παράμετρος $\sigma > 0$ αντιστοιχεί στην τυπική απόκλιση της κατανομής Gauss, η οποία καθορίζει το επίπεδο εξομάλυνσης. Μεγαλύτερες τιμές του σ οδηγούν σε ισχυρότερη καταστολή θορύβου αλλά και σε μεγαλύτερη απώλεια λεπτομερειών.

Στο επόμενο στάδιο υπολογίζεται το μέτρο και η διεύθυνση (γωνία) της κλίσης της εξομαλυμένης εικόνας $f_s(x,y)$, ως:

$$M(x,y) = \sqrt{g_x^2 + g_y^2} \qquad (3.1.18)$$

και

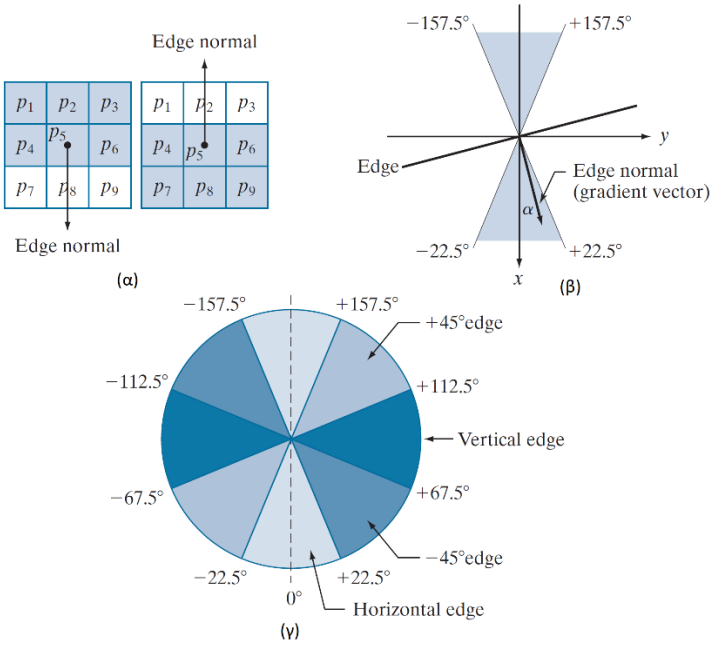
$$a(x,y) = \tan^{-1}\left(\frac{g_y}{g_x}\right) \qquad (3.1.19)$$

Όπου $g_x = \partial f_s/\partial x$ και $g_y = \partial f_s/\partial y$. Οι μερικές παράγωγοι g_x και g_y προσεγγίζονται από οποιαδήποτε ζεύγη μασκών πρώτης τάξης, όπως οι τελεστές Sobel ή Prewitt, για τον εντοπισμό των τοπικών μεταβολών της έντασης. Οι συναρτήσεις $M(x,y)$ και $a(x,y)$ έχουν το ίδιο μέγεθος με την εικόνα $f_s(x,y)$.

Το επόμενο βήμα στον αλγόριθμο Canny είναι η εφαρμογή της εξάλειψη μη μέγιστων (non-maxima suppression). Σκοπός αυτής της διαδικασίας είναι να διατηρηθούν μόνο τα ισχυρότερα σημεία ακμής, απομακρύνοντας τα υπόλοιπα. Βασικά, η διαδικασία στοχεύει στον προσδιορισμό της διεύθυνσης κάθετης στην ακμή, δηλαδή του διανύσματος της κλίσης, ώστε να εντοπιστούν τα τοπικά μέγιστα κατά μήκος αυτής της διεύθυνσης. Στην πράξη, για απλούστερο έλεγχο των γειτονικών εικονοστοιχείων (π.χ. γειτονιά διαστάσεων $3\ x\ 3$), η εξάλειψη μη μέγιστων εφαρμόζεται κατά τέσσερις βασικές διευθύνσεις: οριζόντια (0°), κατακόρυφη (90°), διαγώνια +45° και διαγώνια −45°. Για κάθε εικονοστοιχείο στην θέση (x,y), επιλέγεται μία από τις τέσσερις βασικές διευθύνσεις που είναι η πλησιέστερη της γωνία $a(x,y)$. Στη συνέχεια, ελέγχεται αν το μέτρο της κλίσης $M(x,y)$ αποτελεί τοπικό μέγιστο σε σχέση με τα γειτονικά εικονοστοιχεία κατά τη συγκεκριμένη διεύθυνση εντός της ορισμένης γειτονιά (π.χ. $3\ x\ 3$). Αν δεν πρόκειται για τοπικό μέγιστο, η τιμή του εικονοστοιχείου μηδενίζεται, εξασφαλίζοντας ότι διατηρούνται μόνο τα πιο ισχυρά και σημαντικά σημεία ακμής, διαφορετικά, η τιμή του διατηρείται. Έστω $g_N(x,y)$ είναι η εικόνα με τα εξαλειφθέντα μη μέγιστα και για κάθε εικονοστοιχείο στην θέση (x,y) ορίζεται ως:

$$g_N(x,y) = \begin{cases} M(x,y), & \text{αν το } M(x,y) \text{ είναι τοπικό μέγιστο} \\ & \text{κατά την διεύθυνση της κλισης} \\ 0, & \text{αλλιως} \end{cases} \qquad (3.1.20)$$

Στο σχήμα 3.9 παρουσιάζονται οι περιοχές της γωνίας που αντιστοιχούν στις τέσσερις βασικές διευθύνσεις.



Σχήμα 3.9: (α) Δύο δυνατοί προσανατολισμοί μιας οριζόντιας ακμής, (β) η περιοχή που αντιστοιχεί στη γωνία *α* και (γ) οι περιοχές τιμών για τις τέσσερις βασικές διευθύνσεις (0°, 90°, +45° και −45°) [5]

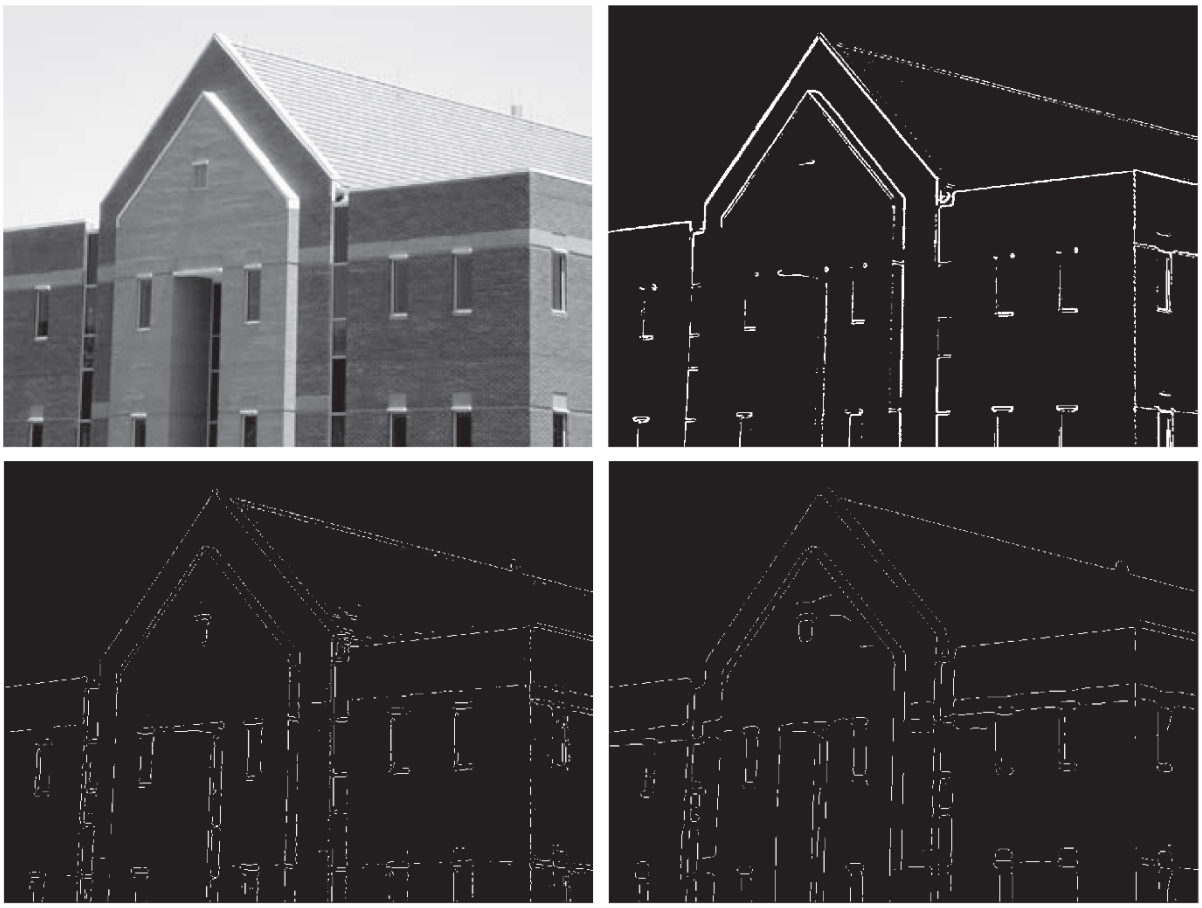
Το τελευταίο βήμα είναι η διπλή κατωφλίωση και η σύνδεση ακμών που εφαρμόζεται στην εικόνα $g_N(x,y)$. Ορίζονται δύο κατώφλια, ένα υψηλό T_{high} και ένα χαμηλό T_{low}. Τα εικονοστοιχεία ταξινομούνται ως εξής:

- **Ισχυρές Ακμές:** $g_N(x,y) \geq T_{high}$
- **Ασθενής Ακμές:** $T_{low} \leq g_N(x,y) < T_{high}$
- **Μηδενικά Στοιχεία:** $g_N(x,y) \leq T_{low}$

Τέλος, εφαρμόζεται σύνδεση ακμών μέσω υστέρησης. Οι ασθενείς ακμές διατηρούνται μόνο εφόσον συνδέονται με ισχυρές ακμές, εξασφαλίζοντας έτσι τη συνέχεια και την αφαίρεση ψευδών ακμών.

Συνοψίζοντας, ο αλγόριθμος Canny για την ανίχνευση ακμών εκτελείται σε τέσσερα βήματα [4,5]:

1. **Εξομάλυνση της εικόνας:** Η εικόνα φιλτράρεται με ένα φίλτρο Gauss για την μείωση του θορύβου.
2. **Υπολογισμός κλίσης:** Υπολογίζονται οι εικόνες του μέτρου και της γωνίας της κλίσης, οι οποίες δείχνουν την ένταση και τη διεύθυνση των αλλαγών έντασης.
3. **Εξάλειψη μη μέγιστων:** Η εικόνα του μέτρου της κλίσης υποβάλλεται σε επεξεργασία ώστε να διατηρούνται μόνο τα τοπικά μέγιστα, δηλαδή τα πιθανά σημεία ακμών.
4. **Διπλή κατωφλίωση και σύνδεση ακμών:** Εφαρμόζονται δύο επίπεδα κατωφλίου και αναλύεται η συνδεσιμότητα των ακμών για την τελική ανίχνευση και σύνδεση των ορίων.



Σχήμα 3. 10: Η αρχική εικόνα με τιμές έντασης από 0 έως 1 (πάνω αριστερά), η κατωφλιωμένη κλίση της εξομαλυμένης εικόνας (πάνω δεξιά), η ανίχνευση ακμών με τον αλγόριθμο LoG (κάτω αριστερά) και η ανίχνευση ακμών με τον αλγόριθμο Canny (κάτω δεξιά) [5]

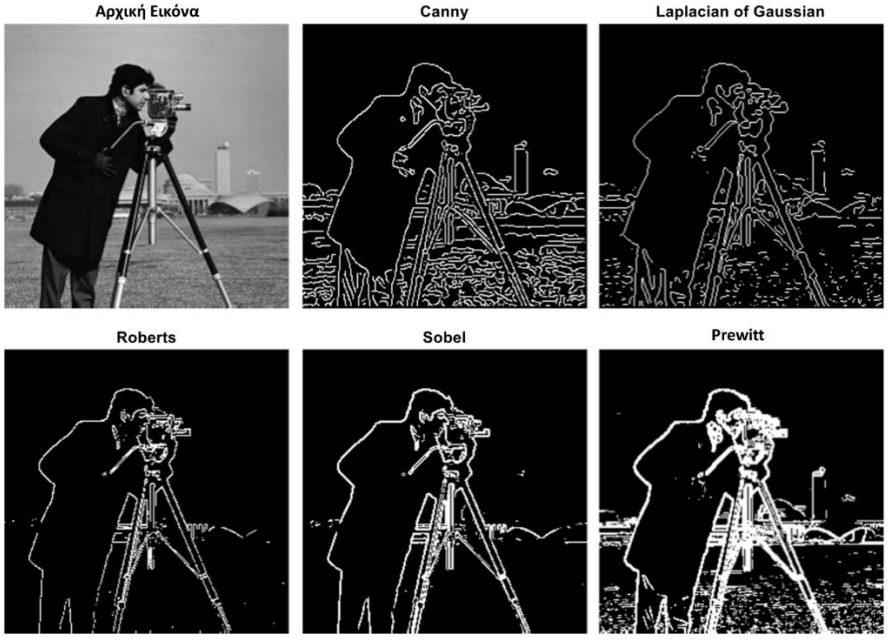
3.1.2.8 Συμπεράσματα

Σε αυτό το κεφάλαιο αναφερθήκαμε στις παραδοσιακές μεθόδους (Roberts, Prewitt, Sobel, LoG και Canny). Οι μέθοδοι αυτές εξακολουθούν να χρησιμοποιούνται σε εφαρμογές όπου απαιτείται γρήγορη και απλή ανίχνευση ακμών. Επιπλέον, παραμένουν ιδιαίτερα χρήσιμες για διδακτικούς σκοπούς, καθώς απεικονίζουν με σαφήνεια τις βασικές αρχές της ανάλυσης κλίσης και της συνέλιξης. Η σύγκριση των παραδοσιακών μεθόδων παρουσιάζεται στον παρακάτω πίνακα, ενώ οι οπτικές διαφορές απεικονίζονται στο σχήμα 3.11.

Σε σύγκριση με σύγχρονες μεθόδους τεχνητής νοημοσύνης και ειδικότερα βαθιάς μάθησης, όπως είναι τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs), οι παροδικές μέθοδοί υστερούν σε ευελιξία και προσαρμοστικότητα. Τέτοιου είδους μέθοδοι τεχνητής νοημοσύνης μπορούν να μάθουν σύνθετα μοτίβα ακμών, να αντιμετωπίζουν διαφορετικά επίπεδα θορύβου και φωτισμού, και να επιτυγχάνουν καλύτερη γενίκευση σε διαφορετικά σύνολα δεδομένων. Αντίθετα, οι παραδοσιακοί τελεστές έχουν περιορισμένη απόδοση σε περιπτώσεις εικόνων με χαμηλή ποιότητα, έντονες σκιάσεις ή περίπλοκες υφές, καθώς βασίζονται σε σταθερούς πυρήνες συνέλιξης και όχι σε δυναμική μάθηση χαρακτηριστικών. Με άλλα λόγια, τα CNNs προσαρμόζονται αυτόματα σε πολύπλοκα μοτίβα και συνθήκες, ενώ οι κλασικές μέθοδοι είναι πιο άκαμπτες, δυσκολεύονται να ανταποκριθούν σε πολύπλοκες ή μεταβαλλόμενες συνθήκες εικόνας και είναι πιο ευάλωτες στον θόρυβο.

Τελεστές	Περιγραφή	Πλεονεκτήματα	Μειονεκτήματα
Robert	Χρήση πρώτης παράγωγου με μικρό πυρήνα 2x2 για γρήγορη ανίχνευση ακμών	Πολύ απλός και γρήγορος	Πολύ ευαίσθητος στο θόρυβο, χαμηλή ακρίβεια θέσης
Prewitt	Πρώτης παράγωγου τελεστής με μεγαλύτερο πυρήνα 3x3 για προσδιορισμό ακμών	Καλύτερος από τον Robert σε ανίχνευση κατευθύνσεων	Περιορισμένη ακρίβεια, ευαισθησία σε θόρυβο
Sobel	Όμοιος με Prewitt, με βάρη που δίνουν μεγαλύτερη έμφαση στο κέντρο για καλύτερο φιλτράρισμα	Καλύτερη ανθεκτικότητα στον θόρυβο σε σχέση με Prewitt	Περιορισμένη ακρίβεια εντοπισμού, μέτρια ευαισθησία στο θόρυβο
LoG	Συνδυασμός Laplacian και Gaussian φιλτραρίσματος για ανίχνευση ακμών	Καταστολή θορύβου πριν την ανίχνευση	Περίπλοκος υπολογισμός, κάποια ευαισθησία στο θόρυβο
Canny	Πολυσταδιακός αλγόριθμος: βελτιστοποίηση εύρεσης ακμών, καταστολή θορύβου και εντοπισμό ενώσεων	Ακριβής, αντιστάθμιση θορύβου, εντοπισμός αδύναμων ακμών	Μεγαλύτερη πολυπλοκότητα, υψηλότερο υπολογιστικό κόστος

Πίνακας 3. 1: Σύγκριση Τελεστών Ανίχνευσης Ακμών.



Σχήμα 3.11: Οπτική σύγκριση των αποτελεσμάτων ανίχνευσης ακμών με τους αλγορίθμους Canny και LoG, καθώς και με τους τελεστές Roberts, Prewitt και Sobel [7]

3.1.3 Τμηματοποίηση Εικόνας

3.1.3.1 Κατηγορίες και Βήματα Τμηματοποίησης

Η τμηματοποίηση εικόνας (image segmentation) είναι η διαδικασία διαχωρισμού μιας εικόνας σε πολλαπλές περιοχές ή τμήματα. Κάθε τμήμα αντιστοιχεί συνήθως σε αντικείμενα, όρια ή περιοχές ενδιαφέροντος στην εικόνα. Αυτή η διαδικασία αποτελεί τη βάση για πολλές εφαρμογές υπολογιστικής όρασης, όπως η αναγνώριση αντικειμένων, η ανίχνευση σκηνών και η επεξεργασία βίντεο. Με τον διαχωρισμό της εικόνας σε τμήματα, καθίσταται δυνατή η ακριβέστερη επεξεργασία κάθε αντικειμένου ξεχωριστά, η ανάλυση των σχέσεων μεταξύ τους, αλλά και η ανάπτυξη πιο προηγμένων τεχνικών υπολογιστικής όρασης. Αυτή η προσέγγιση διευκολύνει την κατανόηση πολύπλοκων σκηνών και βελτιώνει την απόδοση σε εφαρμογές όπως η αυτόνομη οδήγηση, η ιατρική διάγνωση μέσω εικόνων και η ρομποτική.

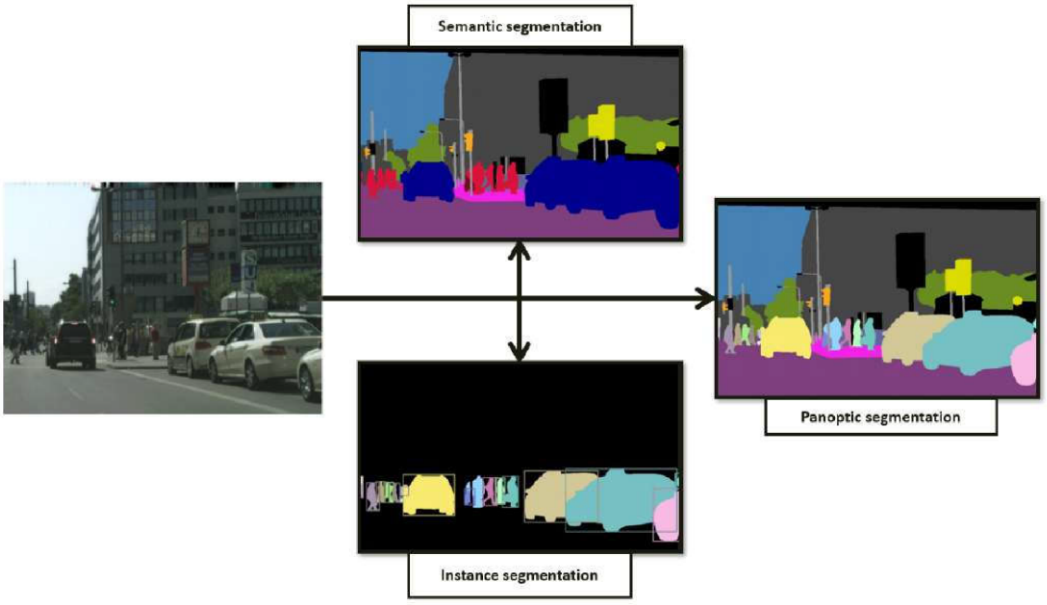
Υπάρχουν τρεις βασικές κατηγορίες τμηματοποίησης [8,9]:

- **Σημασιολογική Τμηματοποίηση (Semantic Segmentation):** Κάθε εικονοστοιχείο της εικόνας κατηγοριοποιείται σε μια κλάση αντικειμένου ή σκηνής (π.χ. "δρόμος", "αυτοκίνητο", "δέντρο"), χωρίς να διαχωρίζονται μεμονωμένα αντικείμενα της ίδιας κατηγορίας (π.χ. δεν κατηγοριοποιεί τον κάθε άνθρωπο ξεχωριστά). Το αποτέλεσμα είναι ένας χάρτης κλάσεων, όπου όλα τα αντικείμενα που ανήκουν στην ίδια κλάση έχουν κοινή ετικέτα.

- **Τμηματοποίηση Παρουσίας (Instance Segmentation):** Αποτελεί ένα περαιτέρω βήμα από τη σημασιολογική τμηματοποίηση, καθώς στοχεύει στην ανίχνευση μεμονωμένων αντικειμένων, ακόμα και αν ανήκουν στην ίδια κλάση. Κάθε αντικείμενο λαμβάνει μια ετικέτα στην τμηματοποιημένη σκηνή, καθώς και ένα πλαίσιο οριοθέτησης γύρω του, ώστε να υποδεικνύεται η τοποθεσία του. Πρόκειται για μια σύνθετη διαδικασία, καθώς το μοντέλο χρειάζεται να διακρίνει με μεγάλη ακρίβεια τις διαφορετικές παρουσίες αντικειμένων που ανήκουν στην ίδια κατηγορία, ξεπερνώντας τόσο τυχόν αλληλοεπικαλύψεις όσο και τις διαφοροποιήσεις στο σχήμα και το μέγεθός τους.

- **Πανοπτική Τμηματοποίηση (Panoptic Segmentation):** είναι μια προηγμένη τεχνική υπολογιστικής όρασης που συνδυάζει τη σημασιολογική και την τμηματοποίηση παρουσίας, με στόχο τη συνολική και ακριβή κατανόηση μιας σκηνής, περιλαμβάνοντας όλες τις κλάσεις και όλα τα μεμονωμένα αντικείμενα. Σε κάθε εικονοστοιχείο εκχωρούνται δύο τιμές: η κατηγορία του, π.χ. "αυτοκίνητο" ή "ουρανός", και ένας αριθμός παρουσίας που ξεχωρίζει τα μεμονωμένα αντικείμενα της ίδιας κατηγορίας (π.χ. «αυτοκίνητο 1» από «αυτοκίνητο 2») .Οι μη μετρήσιμες περιοχές όπως ουρανός ή έδαφος και τα μετρήσιμα αντικείμενα όπως άνθρωποι ή αυτοκίνητα αποδίδονται με μοναδικές μάσκες και χρώματα, εξασφαλίζοντας σαφή διάκριση και ολοκληρωμένη οπτικοποίηση της σκηνής.

Κατανοώντας αυτές τις κατηγορίες, γίνεται σαφές πως η τμηματοποίηση είναι ένα πολυδιάστατο πρόβλημα, που απαιτεί τεχνικές προσαρμοσμένες στη συγκεκριμένη εφαρμογή και τα δεδομένα που εξετάζονται.



Σχήμα 3.12: Οι τρεις βασικές κατηγορίες τμηματοποίησης (Σημασιολογική, Παρουσίας και Πανοπτική) [8]

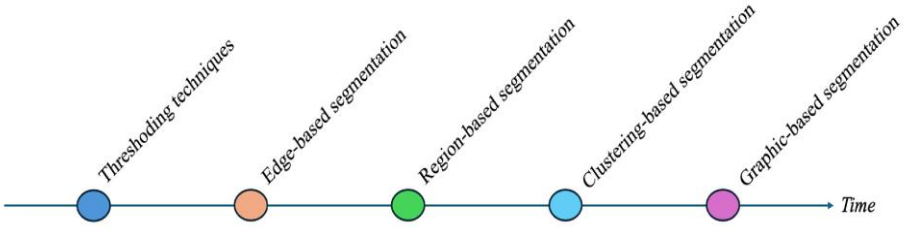
Χαρακτηριστικό	Σημασιολογική Τμηματοποίηση	Τμηματοποίηση Παρουσίας	Πανοπτική Τμηματοποίηση
Σκοπός	Κατηγοριοποίηση κάθε εικονοστοιχείο σε μια κλάση (π.χ. αυτοκίνητο, δρόμος)	Διαχωρισμός και απομόνωση κάθε αντικειμένου της ίδιας κλάσης	Συνδυασμός σημασιολογικής και τμηματικής: ετικέτα κλάσης και κωδικός αντικειμένου για κάθε εικονοστοιχείο
Αναγνώριση Αντικειμένων	Όλα τα αντικείμενα της ίδιας κλάσης θεωρούνται ένα σύνολο	Κάθε αντικείμενο ξεχωριστής παρουσίας έχει μοναδικό κωδικό	Σε κάθε αντικείμενο εκχωρούνται δύο τιμές: η κατηγορία του και ένας αριθμός παρουσίας
Διαχείριση πολλών αντικειμένων	Δεν διαχωρίζει αντικείμενα ίδιας κατηγορίας που επικαλύπτονται	Διαχωρίζει επικάλυψη αντικειμένων με μοναδικές μάσκες	Διαχειρίζεται πλήρως και αντικείμενα και περιοχές
Λεπτομέρεια	Κατηγοριών (γενική)	Αντικειμένων (λεπτομερής)	Πλήρης κατανόηση εικόνας σε επίπεδο κατηγορίας και αντικειμένου

Πίνακας 3.2: Σύγκριση των τριών βασικών κατηγοριών τμηματοποίησης (Σημασιολογική, Παρουσίας και Πανοπτική).

3.1.3.2 Κλασικές Προσεγγίσεις

Για πολλά χρόνια, οι παραδοσιακές τεχνικές τμηματοποίησης αποτελούσαν τη βάση της ανάλυσης εικόνων [10-12], παρέχοντας ποικίλες μεθόδους με εφαρμογές σε ιατρικές, βιομηχανικές και γενικές οπτικές εφαρμογές. Τεχνικές όπως η κατωφλίωση, η ανίχνευση ακμών, η τμηματοποίηση με βάση περιοχές, οι μέθοδοι ομαδοποίησης και οι μέθοδοι γραφημάτων έχουν αποδειχθεί αποτελεσματικές σε πολλές περιπτώσεις. Ωστόσο, συχνά εμφανίζουν σημαντικούς περιορισμούς στην αντιμετώπιση

σύνθετων σεναρίων, όπως εικόνων με παρουσία θορύβου, ασαφή όρια αντικειμένων και πολύπλοκες υφές. Επιπλέον, η έλλειψη δυναμικής προσαρμοστικότητας και δυνατότητας μάθησης από τα δεδομένα περιορίζει την απόδοσή τους σε σύγχρονες εφαρμογές με αυξανόμενες απαιτήσεις. Εξαιτίας αυτών των περιορισμών, η εφαρμογή βαθιών νευρωνικών δικτύων, έχει αναδειχθεί ως η κυρίαρχη μεθοδολογία, καθώς υπερβαίνει τους παραπάνω περιορισμούς, προσδίδοντας αυξημένη ευελιξία και βελτιωμένη ακρίβεια στη διαδικασία της τμηματοποίησης εικόνων.



Σχήμα 3.13: Η χρονική εξέλιξη των παραδοσιακών μεθόδων τμηματοποίησης [10]

Ακολουθεί μια συνοπτική αναφορά στις κλασικές προσεγγίσεις [10-12]:

- **Κατωφλίωση (Thresholding):**

Η κατωφλίωση αποτελεί μια από τις πιο βασικές και διαδεδομένες τεχνικές τμηματοποίησης εικόνας. Ο κύριος σκοπός της είναι η μετατροπή μιας ασπρόμαυρη εικόνας σε δυαδική μορφή, όπου κάθε εικονοστοιχείο θα ταξινομείται σε μία από δύο κατηγορίες, συνήθως «αντικείμενο» ή «φόντο». Η μέθοδος μπορεί να εφαρμοστεί είτε με ένα ενιαίο κατώφλι για ολόκληρη την εικόνα (Global Thresholding) είτε προσαρμοστικά σε διαφορετικές περιοχές της εικόνας (Local Thresholding), ανάλογα με τις συνθήκες φωτισμού και την ομοιογένεια της εικόνας. Οι μέθοδοι με ενιαίο κατώφλι είναι απλές και αποτελεσματικές όταν υπάρχει σαφής διάκριση ανάμεσα στο αντικείμενο και το φόντο. Στην περίπτωση αυτή, επιλέγεται μια σταθερή τιμή κατωφλίου, και κάθε εικονοστοιχείο που έχει τιμή φωτεινότητας μεγαλύτερη ή ίση με αυτήν θεωρείται ότι ανήκει στο αντικείμενο, ενώ τα υπόλοιπα στο φόντο. Μία πολύ γνωστή μέθοδος είναι η Otsu, η οποία επιλέγει το βέλτιστο κατώφλι αυτόματα μεγιστοποιώντας τη διαφορά ανάμεσα στις δύο περιοχές της εικόνας, δηλαδή το αντικείμενο και το φόντο, βασιζόμενη στη στατιστική κατανομή των τιμών φωτεινότητας. Ωστόσο, όταν οι συνθήκες φωτισμού ποικίλλουν ή το φόντο δεν είναι ομοιογενές, οι μέθοδοι με ενιαίο κατώφλι μπορεί να αποτύχουν. Για να αντιμετωπιστούν αυτές οι περιπτώσεις, χρησιμοποιούνται προσαρμοστικές μέθοδοι κατωφλίωσης (Local Thresholding), όπου η τιμή κατωφλίου υπολογίζεται ξεχωριστά για κάθε περιοχή της εικόνας, λαμβάνοντας υπόψη τοπικά χαρακτηριστικά όπως η μέση τιμή και η διασπορά φωτεινότητας. Έτσι, η τμηματοποίηση γίνεται πιο ευαίσθητη στις μεταβολές της εικόνας, επιτρέποντας ακριβέστερο διαχωρισμό αντικειμένου και φόντου σε σύνθετες καταστάσεις.

- **Κατάτμηση με βάση τις ακμές (Edge-Based):**

Η κατάτμηση με βάση τις ακμές στηρίζεται στην ανίχνευση των ορίων μεταξύ ομοιογενών περιοχών μιας εικόνας, όπου παρατηρούνται απότομες μεταβολές στην ένταση των εικονοστοιχείων. Η διαδικασία ανίχνευσης ακμών πραγματοποιείται μέσω κατάλληλων φίλτρων, τα οποία αναδεικνύουν τις εντονότερες μεταβολές της έντασης, καθιστώντας δυνατή την αναγνώριση των περιγραμμάτων των αντικειμένων και τον προσδιορισμό των ορίων τους. Η μέθοδος αυτή αποδίδει ιδιαίτερα αποτελεσματικά όταν τα αντικείμενα στην εικόνα έχουν σαφή και καλά καθορισμένα όρια, ενώ η χρήση φίλτρων εξομάλυνσης βοηθά στη μείωση του θορύβου πριν από την ανίχνευση ακμών, βελτιώνοντας την ακρίβεια της τμηματοποίησης. Μια από τις πλέον χρησιμοποιούμενες τεχνικές σε αυτό το πλαίσιο είναι ο αλγόριθμος Canny (βλέπε Ενότητα 3.1.2.7 Ο ανιχνευτής Ακμών Canny, ο

οποίος συνδυάζει υψηλή ακρίβεια στον εντοπισμό ακμών με αποτελεσματική καταστολή θορύβου, καθιστώντας αξιόπιστο για ποικίλες εφαρμογές ανάλυσης εικόνας.

- **Κατάτμηση με βάση τις περιοχές (Region-Based):**

Η κατάτμηση με βάση τις περιοχές στηρίζεται στην υπόθεση ότι τα εικονοστοιχεία που ανήκουν στην ίδια περιοχή μοιράζονται κοινά χαρακτηριστικά, όπως χρώμα, φωτεινότητα ή υφή, ενώ ταυτόχρονα βρίσκονται σε γειτονικές θέσεις. Με αυτόν τον τρόπο, η εικόνα χωρίζεται σε ομάδες εικονοστοιχείων με κοινές ιδιότητες, γεγονός που επιτρέπει πιο λεπτομερή και ουσιαστική ανάλυση του οπτικού περιεχομένου της. Δύο από τις πιο αντιπροσωπευτικές προσεγγίσεις αυτής της κατηγορίας είναι η Seeded Region Growing (SRG), η οποία ξεκινά από προκαθορισμένα σημεία-«σπόρους» και επεκτείνει σταδιακά τις περιοχές με βάση προκαθορισμένα κριτήρια ομοιότητας, και η Split-and-Merge, η οποία στηρίζεται σε μια ιεραρχική διαδικασία διάσπασης της εικόνας σε μικρότερες περιοχές και στη συνέχεια συγχώνευσής τους, ανάλογα με την ομοιογένεια που παρουσιάζουν. Οι μέθοδοι αυτές είναι ιδιαίτερα χρήσιμες όταν οι περιοχές ενδιαφέροντος παρουσιάζουν σαφή και ομοιόμορφα χαρακτηριστικά, διευκολύνοντας τον ακριβέστερο διαχωρισμό αντικειμένων.

- **Μέθοδοι Ομαδοποίησης (Clustering-Based):**

Οι μέθοδοι ομαδοποίησης βασίζονται στην ιδέα ότι τα εικονοστοιχεία μιας εικόνας μπορούν να ταξινομηθούν σε ομάδες (clusters), έτσι ώστε τα εικονοστοιχεία κάθε ομάδας να μοιράζονται περισσότερα κοινά χαρακτηριστικά μεταξύ τους σε σύγκριση με τις άλλες ομάδες. Με αυτόν τον τρόπο, η εικόνα διαχωρίζεται σε τμήματα με κοινά χαρακτηριστικά, διευκολύνοντας την αναγνώριση και ανάλυση των αντικειμένων. Οι μέθοδοι ομαδοποίησης διακρίνονται σε δύο κύριες κατηγορίες. Στην πρώτη κατηγορία ανήκουν οι ιεραρχικές μέθοδοι (hierarchical clustering), οι οποίες δημιουργούν μια δενδροειδή δομή (dendrogram) μέσω μιας σταδιακής διαδικασίας συγχώνευσης (agglomerative hierarchical clustering) ή διάσπασης (divisive hierarchical clustering) των ομάδων. Στη δεύτερη κατηγορία εντάσσονται οι διαμεριστικές μέθοδοι (partitional clustering), όπως οι αλγόριθμοι k-means και Fuzzy c-means, οι οποίες κατατάσσουν δεδομένα σε διακριτές ομάδες μέσω της βελτιστοποίησης ενός κριτηρίου αξιολόγησης, ώστε η ομοιότητα εντός κάθε ομάδας να είναι μεγαλύτερη σε σύγκριση με τις άλλες. Παρότι οι μέθοδοι ομαδοποίησης προσφέρουν ευελιξία μπορούν να εφαρμοστούν σε ποικίλες εικόνες, παρουσιάζουν περιορισμούς, όπως ευαισθησία στον θόρυβο και εξάρτηση από τον προκαθορισμό του αριθμού των ομάδων.

- **Μέθοδοι Γραφημάτων (Graph-Based):**

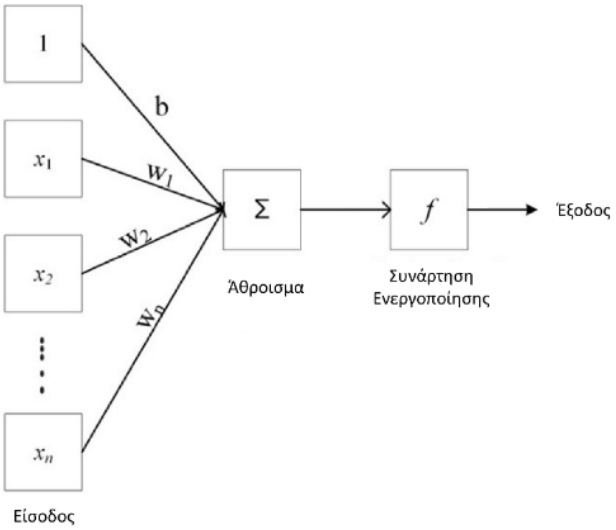
Οι μέθοδοι αυτές αναπαριστούν την εικόνα ως γράφο, όπου τα εικονοστοιχεία ή περιοχές της εικόνας αντιστοιχούν σε κόμβους, ενώ οι ακμές εκφράζουν σχέσεις γειτνίασης ή ομοιότητας μεταξύ τους. Η τμηματοποίηση προκύπτει με τον χωρισμό της εικόνας σε υπογράφους, καθένας από τους οποίους αποτελείται από διακριτά και μη επικαλυπτόμενα τμήματα με κοινά χαρακτηριστικά, ενώ η ένωση όλων αυτών καλύπτει πλήρως το αρχικό σύνολο κόμβων. Μερικές από τις κύριες μεθοδολογίες τμηματοποίησης εικόνων που βασίζονται στη χρήση γράφων περιλαμβάνουν: τα Minimal Spanning Tree (MST) που βασίζονται στην ιδέα ότι η εικόνα μπορεί να αναπαρασταθεί ως ένα δέντρο που ενώνει όλους τους κόμβους με το μικρότερο δυνατό συνολικό βάρος, διευκολύνοντας τον εντοπισμό ομοιογενών περιοχών. Οι Graph Cuts, που προσεγγίζουν την τμηματοποίηση ως πρόβλημα βέλτιστης τομής στον γράφο, διαχωρίζοντας τα εικονοστοιχεία σε αντικείμενα και φόντο με βάση προκαθορισμένα κριτήρια. Τα Markov Random Fields (MRF) εφαρμόζουν πιθανοτικά μοντέλα για την κατηγοριοποίηση των εικονοστοιχείων, αξιοποιώντας τα χαρακτηριστικά των γειτονικών εικονοστοιχείων ώστε να επιτευχθεί πιο ακριβής τμηματοποίηση. Ένα από τα κύρια πλεονεκτήματα είναι η δυνατότητά τους να αξιοποιούν εκ των προτέρων πληροφορία για τη βελτίωση των αποτελεσμάτων. Τέλος, οι μέθοδοι συντομότερης διαδρομής (Shortest Path) βασίζονται στον

εντοπισμό της διαδρομής μεταξύ δύο κόμβων σε έναν γράφο, η οποία ελαχιστοποιεί το συνολικό άθροισμα των βαρών των ακμών που τη συνθέτουν. Ως αποτέλεσμα, κάθε διαδρομή αντιστοιχεί σε μια αλληλουχία κόμβων που μοιράζονται κοινά χαρακτηριστικά, καθορίζοντας έτσι τις διαφορετικές περιοχές της τμηματοποίησης.

3.1.3.3 Μοντέρνες Προσεγγίσεις

Σε αυτήν την ενότητα παρουσιάζονται οι προσεγγίσεις με χρήση βαθιάς μάθησης (ΒΜ), η οποία αποτελεί μια υποκατηγορία της τεχνητής νοημοσύνης. Η ΒΜ επιτρέπει την αυτόματη εξαγωγή χαρακτηριστικών από τα δεδομένα, περιορίζοντας την ανάγκη για ανθρώπινη παρέμβαση, όπως η προεπεξεργασία των δεδομένων ή η παροχή εξειδικευμένης γνώσης για κάθε εφαρμογή. Επιπλέον, τα βαθιά νευρωνικά δίκτυα μπορούν να αναπαραστήσουν πολύπλοκες, μη γραμμικές σχέσεις, επιτυγχάνοντας υψηλή ακρίβεια σε εφαρμογές όπως αναγνώριση εικόνας, επεξεργασία φυσικής γλώσσας και άλλες χρήσεις της τεχνητής νοημοσύνη. Για να κατανοηθούν οι βασικές αρχές της ΒΜ, είναι σημαντικό να αναφερθούμε στα τεχνητά νευρωνικά δίκτυα (ΤΝΔ, Artificial Neural Networks) [13-15], τα οποία αποτελούν τη θεμελιώδη δομή πάνω στην οποία στηρίζονται οι περισσότερες σύγχρονες μέθοδοι.

Ένας τεχνητός νευρώνας (Σχήμα 3.14) βασίζεται στον τρόπο λειτουργίας των νευρώνων του εγκεφάλου και αποτελεί το βασικό δομικό στοιχείο ενός τεχνητού νευρωνικού δικτύου [13-15]. Δέχεται ως είσοδο τις αριθμητικές τιμές $x_1, x_2, \dots, x_n$, οι οποίες πολλαπλασιάζονται με τα αντίστοιχα βάρη $w_1, w_2, \dots, w_n$. Στη συνέχεια, υπολογίζει το σταθμισμένο άθροισμα των εισόδων, στο οποίο προστίθεται ένας όρο πόλωσης (bias) b. Ο όρος αυτός λειτουργεί ως σταθερή μετατόπιση, επιτρέποντας στον νευρώνα να ενεργοποιείται ακόμη και όταν οι είσοδοι είναι μηδενικές ή χαμηλές. Τέλος, εφαρμόζεται μια συνάρτηση ενεργοποίησης (activation function) f, η οποία αποτελεί μη γραμμική συνάρτηση που μετατρέπει το σταθμισμένο άθροισμα σε τελική απόκριση. Με τον τρόπο αυτό, το νευρωνικό δίκτυο μπορεί να μαθαίνει και να αναπαριστά πολύπλοκα μοτίβα. Η απόκριση-έξοδος του τεχνητού νευρώνα μπορεί είτε να τροφοδοτήσει άλλους νευρώνες σε επόμενα επίπεδα, είτε να αποτελέσει το τελικό αποτέλεσμα του δικτύου.



Σχήμα 3.14: Μοντέλο τεχνητού νευρώνας [15]

Μερικές από τις πιο ευρέως χρησιμοποιούμενες συναρτήσεις ενεργοποίησης περιλαμβάνουν [13-15]:

- **Σιγμοειδή συνάρτηση - Sigmoid:** Έχουν την μορφή τελικού σίγμα και Μετατρέπει οποιαδήποτε είσοδο σε τιμές μεταξύ 0 και 1, κάτι που την καθιστά χρήσιμη για την αναπαράσταση πιθανοτήτων.

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

- **Υπερβολική εφαπτομένη – Tanh:** Παρόμοια με τη Sigmoid, αλλά επιστρέφει τιμές στο εύρος [−1,1], προσφέροντας πιο ισορροπημένη έξοδο γύρω από το μηδέν.

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

- **Rectified Linear Unit (ReLU):** Επιστρέφει 0 για αρνητικές τιμές και την ίδια την τιμή για θετικές εισόδους. Η απλότητά της την καθιστά ιδιαίτερα αποτελεσματική στην εκπαίδευση βαθιών νευρωνικών δικτύων.

$$ReLU(x) = \max(0, x)$$

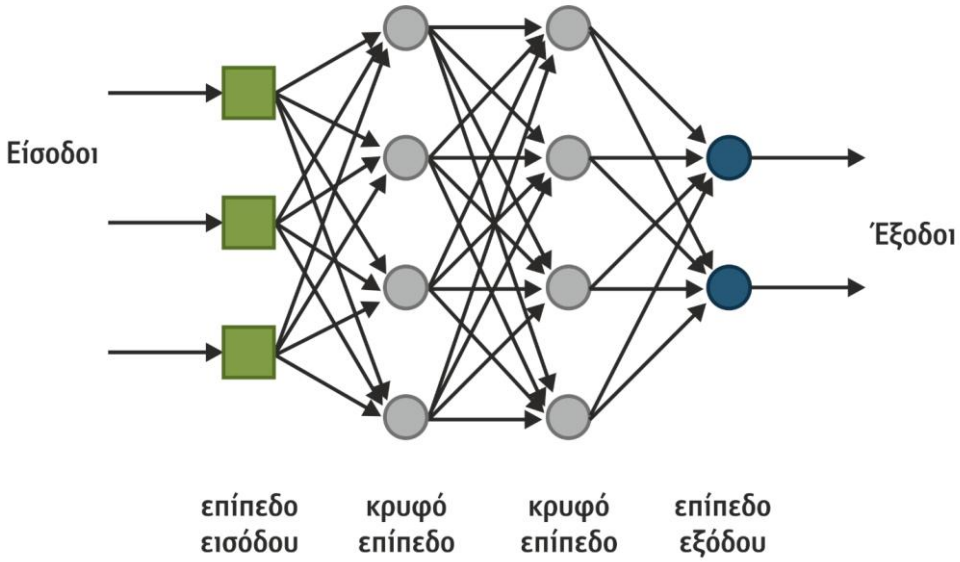
Μια από τις πιο γνωστές αρχιτεκτονικές νευρωνικών δικτύων είναι τα δίκτυα πρόσθιας τροφοδότησης (Feedforward Neural Networks), τα οποία περιλαμβάνουν τα εξής επίπεδα [13-14]:

- **Επίπεδο εισόδου (Input Layer):** αποτελεί το πρώτο επίπεδου του δικτύου και λαμβάνει τα αρχικά δεδομένα. Κάθε νευρώνας αυτού του επιπέδου αντιστοιχεί σε ένα χαρακτηριστικό των δεδομένων εισόδου (π.χ. λέξεις σε κείμενο ή μετρήσεις αισθητήρων).

- **Κρυφά επίπεδα (Hidden Layers):** αποτελείτε από ένα ή περισσότερα επίπεδα που τοποθετούνται μεταξύ εισόδου και εξόδου. Είναι υπεύθυνα για την εκμάθηση πολύπλοκων μοτίβων στα δεδομένα.

- **Επίπεδο εξόδου (Output Layer):** παράγει την τελική απόκριση ή πρόβλεψη. Ο αριθμός των νευρώνων σε αυτό το επίπεδο αντιστοιχεί είτε στον αριθμό των κλάσεων σε ένα πρόβλημα ταξινόμησης (π.χ. η θερμοκρασία μπορεί να θεωρηθεί ως «χαμηλή», «μέση» ή «υψηλή»), είτε στον αριθμό των τιμών προς πρόβλεψη σε ένα πρόβλημα παλινδρόμησης (π.χ. η θερμοκρασία αύριο).

Τα δίκτυα με πολλαπλά κρυφά επίπεδα, γνωστά ως βαθιά νευρωνικά δίκτυα (Deep Neural Networks), έχουν τη δυνατότητα να αποτυπώσουν πολύπλοκες σχέσεις μέσα στα δεδομένα.



Σχήμα 3.15: Ένα παράδειγμα δικτύου[1] πρόσθιας τροφοδότησης

Η εκπαίδευση του νευρωνικού δικτύου πραγματοποιείται μέσω της τροποποίησης των βαρών των τεχνητών νευρώνων, ώστε η έξοδος του δικτύου να πλησιάζει όσο το δυνατόν περισσότερο στα επιθυμητά αποτελέσματα. Η διαδικασία αυτή συνήθως βασίζεται σε έναν αλγόριθμο ελαχιστοποίησης σφάλματος, όπως η μέθοδος της αναστροφής μετάδοσης λάθους (back-propagation), κατά την οποία υπολογίζεται το σφάλμα μεταξύ της πραγματικής και της προβλεπόμενης εξόδου και προωθείται αντίστροφα στο δίκτυο, καθορίζοντας σταδιακά πώς το κάθε βάρος συνέβαλε στο συνολικό σφάλμα. Στη συνέχεια, τα βάρη ενημερώνονται χρησιμοποιώντας τεχνικές βελτιστοποίησης, ώστε να μειωθεί το σφάλμα στο επόμενο βήμα.
Ακολουθεί μια συνοπτική αναφορά στις μοντέρνες προσεγγίσεις που υλοποιούνται σε εικόνες [10-17]:

- **Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs):**

Τα CNNs [10,14,18-19] αποτελούν μια από τις ευρέως χρησιμοποιούμενες αρχιτεκτονικές για την επεξεργασία εικόνων και την αναγνώριση οπτικών μοτίβων. Τα κύρια πλεονεκτήματα τους είναι ότι δέχονται ως είσοδο ολόκληρη την εικόνα, χωρίς περεταίρω επεξεργασία της και μπορούν να εξάγουν αυτόματα τα σημαντικά χαρακτηριστικά, χωρίς την ανθρώπινη παρέμβαση

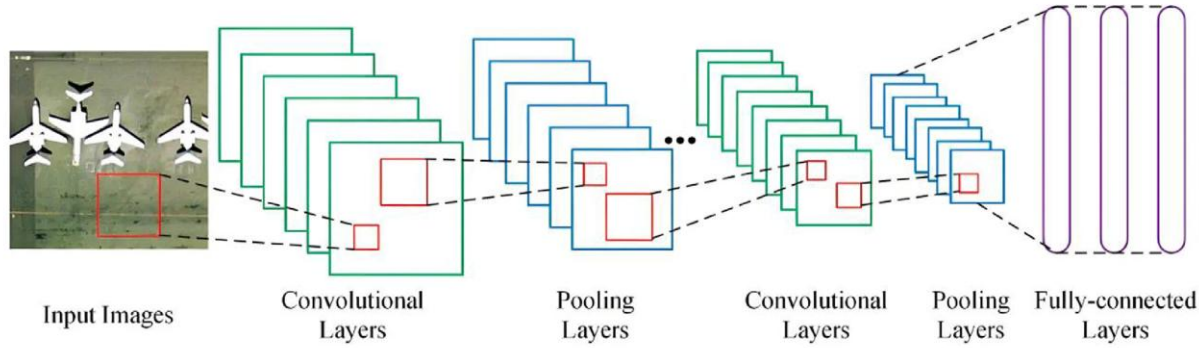
Η αρχιτεκτονική ενός CNN περιλαμβάνει τρία βασικά είδη επιπέδων:

- **Συνελικτικά επίπεδα (Convolutional Layers):** Εξάγουν τα σημαντικά χαρακτηριστικά της εικόνα, όπως ακμές, γωνίες και υφές, εφαρμόζοντας μάσκες (π.χ. διαστάσεων $3\ x\ 3$, όπως αναλύθηκαν στο κεφάλαιο αναγνώρισης ακμής) σε τοπικές περιοχές της εικόνας μέσω της διαδικασίας της συνέλιξης.

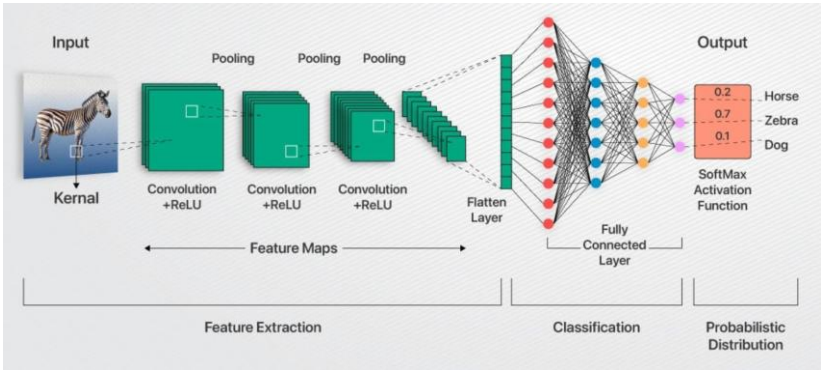
- **Συγκεντρωτικά επίπεδα (Pooling Layers):** Το συγκεντρωτικό επίπεδο εφαρμόζεται στα αποτελέσματα του συνελικτικού επιπέδου, χρησιμοποιώντας τεχνικές όπως επιλογή μεγαλύτερης (max pooling) ή μέση (average pooling) τιμής μέσα σε ένα μικρό παράθυρο τιμών (π.χ. διαστάσεων $2\ x\ 2$), μειώνοντας έτσι τις διαστάσεις και διατηρώντας τις πιο σημαντικές πληροφορίες.

- **Πλήρως συνδεδεμένα επίπεδα (Fully Connected Layers):** Λαμβάνουν τους χάρτες χαρακτηριστικών (feature maps) και παράγουν τις τελικές προβλέψεις, χρησιμοποιώντας μη γραμμικές συναρτήσεων ενεργοποίησης, όπως Sigmoid ή ReLU, ανάλογα με την εφαρμογή. Κάθε χάρτης χαρακτηριστικών αποτελεί μια δισδιάστατη αναπαράσταση που αποτυπώνει συγκεκριμένα χαρακτηριστικά ή μοτίβα, όπως ακμές, γωνίες ή πιο πολύπλοκα σχήματα, ανάλογα με το στάδιο του δικτύου, τα οποία ανιχνεύονται από τα φίλτρα κατά τη διαδικασία της συνέλιξης.

Το σχήμα 3.16 απεικονίζει την γενική αρχιτεκτονική ενός CNN. Η αρχική εικόνα διοχετεύεται σε πολλαπλά διαδοχικά συνελικτικά (μπλε) και συγκεντρωτικά (πράσινο) επίπεδα. Η αλληλουχία αυτών των επιπέδων επαναλαμβάνεται αρκετές φορές, επιτρέποντας στο δίκτυο να αναγνωρίζει ολοένα και πιο πολύπλοκα και αφηρημένα χαρακτηριστικά. Τέλος, οι χάρτες χαρακτηριστικών μετατρέπονται σε μονοδιάστατο διάνυσμα (flattening) ώστε να μπορέσουν να περάσουν στα πλήρως συνδεδεμένα επίπεδα του δικτύου (fully connected layers, μωβ), όπου τα χαρακτηριστικά συνδυάζονται για να ληφθεί η τελική απόφαση, συνήθως υπό μορφή ταξινόμησης ή πρόβλεψης. Τα CNNs έχουν χρησιμοποιηθεί σε εφαρμογές όπως ταξινόμηση εικόνων, ανίχνευση αντικειμένων και αναγνώριση προσώπων.



Σχήμα 3.16: Απεικόνιση της γενικής αρχιτεκτονικής ενός CNN [15]



Σχήμα 3.17: Παράδειγμα αναγνώρισης ζέβρας με χρήση CNN[1], με κύρια στάδια την εξαγωγή χαρακτηριστικών (Feature Extraction), την ταξινόμηση (Classification) και την πιθανοτική κατανομή (Probilistic Distribution) της εξόδου.

[1] analytixlabs.co.in/blog/convolutional-neural-network/

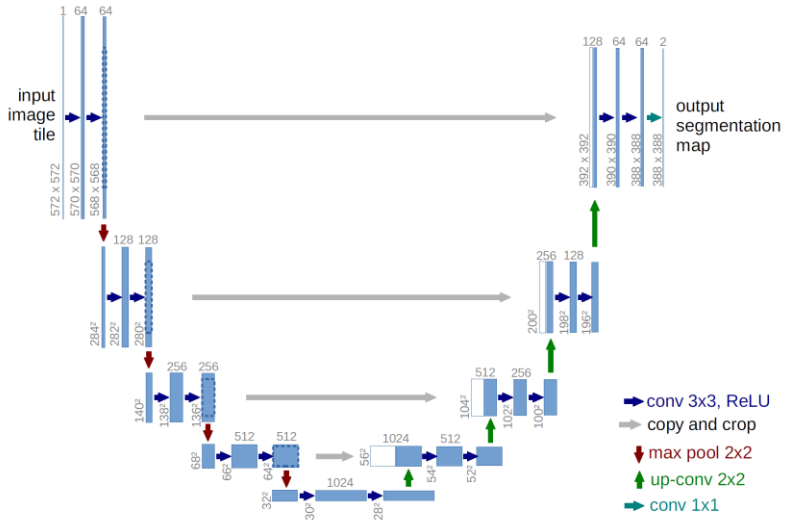
- **Μοντέλο U-Net:**

Το U-Net [10,12,16,17,20] είναι μια από τις πιο δημοφιλείς αρχιτεκτονικές για εφαρμογές σημασιολογικής τμηματοποίησης, λόγω των εξαιρετικών αποδόσεων που επιτυγχάνει. Η αρχιτεκτονική αυτή πήρε το όνομα της από το χαρακτηριστικό σχήμα της, που θυμίζει γράμμα "U", όπως φαίνεται στο σχήμα 3.18. Η αρχιτεκτονική του U-Net αποτελείται από δύο βασικά μέρη:

- Το **συμβατικό μονοπάτι (contracting path),** γνωστό και ως κωδικοποιητής (encoder), απεικονίζεται στην αριστερή πλευρά του σχήματος 3.18. Έχει σχεδιαστεί για την εξαγωγή ιεραρχικών χαρακτηριστικών από την εικόνα εισόδου. Συνήθως αποτελείται από πολλαπλά διαδοχικά τμήματα, όπου σε καθένα εφαρμόζονται δύο συνελίξεις $3\,x\,3$, οι οποίες συχνά ακολουθούνται από συνάρτηση ενεργοποίησης ReLU. Κατά τη μετάβαση προς βαθύτερα επίπεδα του μοντέλου U-Net, πραγματοποιείται υποδειγματοληψία μέσω επιλογή μεγαλύτερης τιμής (max pooling) $2\,x\,2$, προκειμένου να μειωθούν σταδιακά οι χωρικές διαστάσεις των χαρτών χαρακτηριστικών (feature maps). Ένα σημαντικό χαρακτηριστικό του συμβατικού μονοπατιού είναι ότι διπλασιάζεται ο αριθμός των καναλιών χαρακτηριστικών (feature channels) μετά από κάθε βήμα (π.χ. από 64 σε 128, μετά σε 256 κ.ο.κ.), γεγονός που επιτρέπει στο δίκτυο να μαθαίνει ολοένα και πιο αφηρημένες αναπαραστάσεις. Ο αριθμός καναλιών είναι το πλήθος των διαφορετικών χαρτών χαρακτηριστικών.

- Το **επεκτατικό μονοπάτι (expanding path),** γνωστό και ως αποκωδικοποιητής (decoder), απεικονίζεται στην δεξιά πλευρά του σχήματος 3.18. Πραγματοποιεί την αντίστροφη διαδικασία από τον κωδικοποιητή, εφαρμόζοντας υπερδειγματοληψία (upsampling) με συνελίξεις $2\,x\,2$ (up-convolutions). Στόχος του είναι να αυξήσει τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών σε κάθε βήμα, δηλαδή να μεγεθύνει την εικόνα σταδιακά. Ταυτόχρονα, μειώνει τον αριθμό των καναλιών για να περιορίσει την πολυπλοκότητα και να εστιάσει στα πιο σημαντικά χαρακτηριστικά. Κατά τη διαδικασία αυτή, κάθε επίπεδο του αποκωδικοποιητή συγχωνεύεται με τους αντίστοιχους χάρτες χαρακτηριστικών από το συμβατικό μονοπάτι μέσω των συνδέσεων παράκαμψης (skip connections), που απεικονίζονται με γκρι βέλη στο σχήμα 3.18. Αυτές οι συνδέσεις επιτρέπουν τη μεταφορά σημαντικών χαρακτηριστικών υψηλής ανάλυσης από τον κωδικοποιητή στον αποκωδικοποιητή, διατηρώντας τη λεπτομέρεια και βελτιώνοντας την ακρίβεια στον εντοπισμό ορίων.

Το συνδετικό κομμάτι μεταξύ των δύο μονοπατιών του U-Net, γνωστό ως bottleneck, αποτελεί το βαθύτερο επίπεδο του δικτύου. Λαμβάνει από τον κωδικοποιητή την πιο συμπυκνωμένη και αφαιρετική αναπαράσταση των χαρακτηριστικών της εικόνας, την επεξεργάζεται περαιτέρω για την εξαγωγή σύνθετων μοτίβων και τη μεταδίδει στον αποκωδικοποιητή, ώστε να ξεκινήσει η διαδικασία ανακατασκευής της χωρικής ανάλυσης και της λεπτομερειακής τμηματοποίησης.



Σχήμα 3.18: Αρχιτεκτονική του U-Net [20]

Η αρχιτεκτονική του U-Net έχει εξελιχθεί σημαντικά μέσα από πολυάριθμες παραλλαγές και βελτιώσεις [21], γεγονός που την καθιστά σημείο αναφοράς στον τομέα της σύγχρονης τμηματοποίησης εικόνων

- **DeepLab:**

Η οικογένεια αρχιτεκτονικών DeepLab αντιπροσωπεύει μία σημαντική εξέλιξη στη σημασιολογική τμηματοποίηση εικόνων, με κάθε νέα εκδοχή να βασίζεται στην προηγούμενη, αντιμετωπίζοντας προκλήσεις και βελτιώνοντας σταδιακά την απόδοση [11,12,16,17,22,23].

Σημαντικές εκδόσεις της οικογένειας DeepLab:

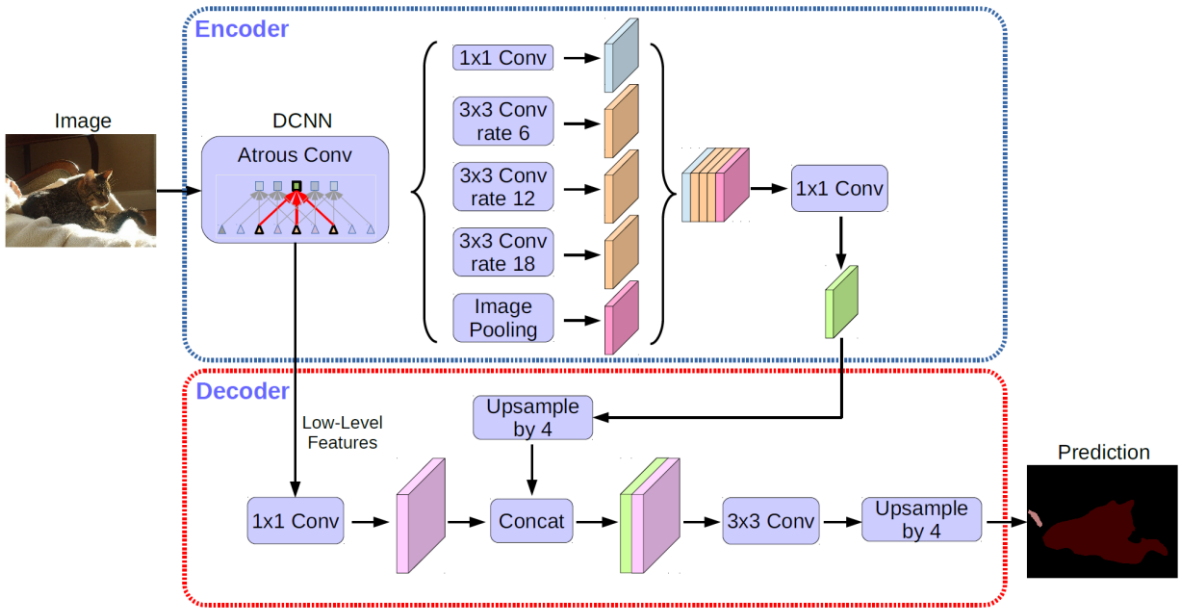
- **DeepLabV1**: Εισήγαγε τα atrous (ή dilated) συνελικτικά επίπεδα και επέτρεψε στο μοντέλο να καταγράψει ένα ευρύτερο φάσμα πληροφορίας μέσα στις εικόνες χωρίς απώλεια ανάλυσης.

- **DeepLabV2**: Επέκτεινε την αρχιτεκτονική με το *Atrous Spatial Pyramid Pooling (ASPP)*, διευκολύνοντας την αναγνώριση αντικειμένων με διαφορετικά μεγέθη.

- **DeepLabV3**: Βελτίωσε περαιτέρω το ASPP και την αρχιτεκτονική για υψηλότερη απόδοση.

- **DeepLabV3+**: Συνδύασε το ASPP με αρχιτεκτονική encoder–decoder, βελτιώνοντας την ανάκτηση λεπτομερειών και την ακρίβεια στα όρια των αντικειμένων. Η αρχιτεκτονική του φαίνεται στο σχήμα 3.19.

Η συνεχής εξέλιξη της οικογένειας DeepLab οδήγησε σε κορυφαίες επιδόσεις και μεγαλύτερη και βελτιωμένη ικανότητα διαχείρισης πολύπλοκων σκηνών.



Σχήμα 3.19: DeepLabV3+ αρχιτεκτονική [23]

- **Vision Transformer (ViT):**

Ο Vision Transformer (ViT) αποτελεί τεχνολογία αιχμής στον τομέα της μηχανικής όρασης και βασίζονται στο μοντέλο Transformer, το οποίο αρχικά αναπτύχθηκε για εφαρμογές επεξεργασίας φυσικής γλώσσας [24–26].

Τα κύρια χαρακτηριστικά του ViT:

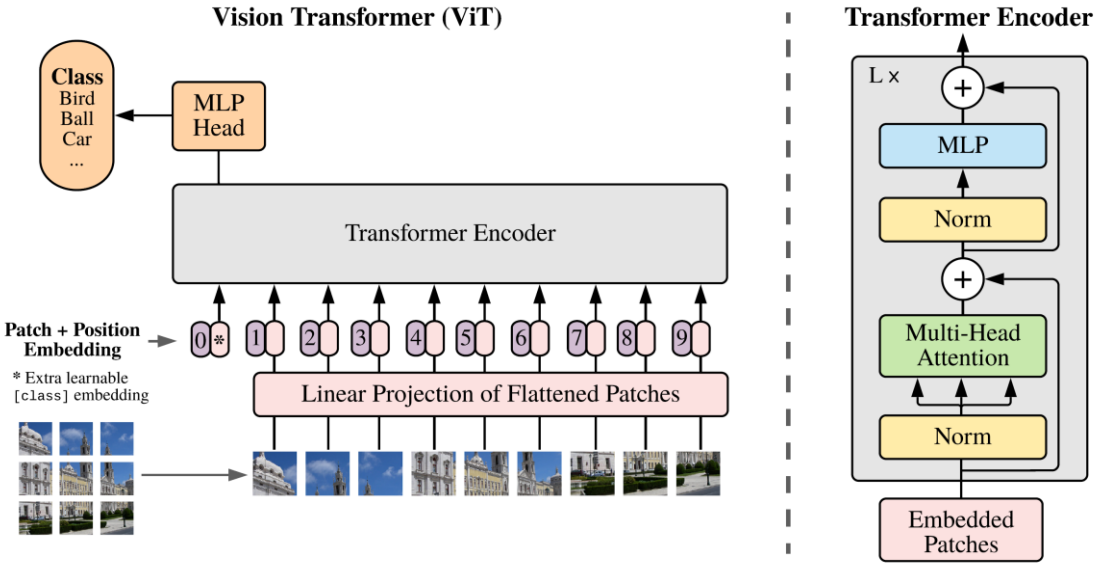
1. **Διαχωρισμός Εικόνας σε Τεμάχια**: Η εικόνα χωρίζεται σε ακολουθία από μικρότερα τεμάχια (patches, π.χ. 16x16 pixels) και κάθε τεμάχιο μετατρέπεται σε ένα διάνυσμα μέσω γραμμικής προβολής.

2. **Προσθήκη Θέσης:** Προκειμένου ο Transformer να διατηρήσει τη χωρική πληροφορία, σε κάθε τεμάχιο προστίθεται διανυσματική αναπαράσταση της θέσης των εικονοστοιχείων.

3. **Επεξεργασία με Transformer Encoder:** Η ακολουθία διανυσμάτων ενσωμάτωσης (embedded patches), που περιλαμβάνει τόσο την πληροφορία περιεχομένου όσο και τη χωρική τους θέση, μαζί με ένα επιπλέον χαρακτηριστικό εκμάθησης κλάσης (learnable class token), εισάγεται σε έναν κωδικοποιητή (Transformer Encoder). Ο κωδικοποιητής αυτός αποτελείται από πολλαπλά επίπεδα μηχανισμού αυτοπροσοχής (self-attention) και πλήρως συνδεδεμένων (feed-forward) επιπέδων, ενισχύοντας τη σταθερότητα και την εκπαίδευση του δικτύου.

4. **Κατηγοριοποίηση (Classification):** Η έξοδος από το τελευταίο επίπεδο του κωδικοποιητή αποτελεί το εκπαιδευμένο διάνυσμα κλάσης (class token), το οποίο λειτουργεί ως αναπαράσταση της εικόνας και τροφοδοτείται σε έναν multi-layer perceptron (MLP), επιτρέποντας στο μοντέλο να προβλέψει την κατηγορία της εικόνας.

Ο ViT προσφέρει σημαντικά πλεονεκτήματα και έχει ευρεία εφαρμογή. Η δυνατότητα αυτοπροσοχής επιτρέπει την καθολική εκμάθηση μακρινών εξαρτήσεων, εντοπίζοντας σχέσεις μεταξύ απομακρυσμένων περιοχών της εικόνας, κάτι που αποτελεί πρόκληση για τα παραδοσιακά CNNs, τα οποία περιορίζονται σε περιοχές που καθορίζονται από τα φίλτρα τους. Ο ViT έχει χρησιμοποιηθεί επιτυχώς σε διάφορες εφαρμογές, όπως κατηγοριοποίηση εικόνας, ανίχνευση αντικειμένων και τμηματοποίηση εικόνας (σε όλες τις κατηγορίες: σημασιολογική, τμηματοποίηση και πανοπτική). Επιπλέον, ο ViT παρουσιάζει υψηλή ανθεκτικότητα σε θόρυβο και παραμορφώσεις εικόνας, διατηρώντας ικανοποιητική απόδοση ακόμη και όταν σημαντικά τμήματα της εικόνας είναι μη ορατά ή αλλοιωμένα. Τέλος, ο ViT εκπαιδεύεται αρχικά σε ένα μεγάλο σύνολο δεδομένων (pretraining) και στη συνέχεια προσαρμόζεται (fine-tuning) σε μικρότερα σύνολα δεδομένων, εξασφαλίζοντας αποτελεσματική γενίκευση σε διαφορετικές εφαρμογές, αν και απαιτεί σημαντικούς υπολογιστικούς πόρους.



Σχήμα 3.20: Vision Transformers αρχιτεκτονική [24]

3.1.4 Εφαρμογές σε Πολιτιστική Κληρονομιά

Η αναγνώριση εικόνας και χρώματος μέσω ψηφιακής επεξεργασίας, ιδίως με την αξιοποίηση αλγορίθμων τεχνητής νοημοσύνης, προσφέρει νέες δυνατότητες στον τομέα της πολιτιστικής κληρονομιάς, διευκολύνοντας τη διατήρηση, την ανάδειξη και την αποκατάσταση των πολιτιστικών αγαθών, διασφαλίζοντας παράλληλα την αυθεντικότητα και την ιστορική ακρίβεια τους [30]. Πιο συγκεκριμένα:

Διατήρηση - Ανάδειξη:

Η σύγχρονη τεχνολογία αναγνώρισης εικόνας και χρώματος προσφέρει νέες δυνατότητες στη διατήρηση της πολιτιστικής κληρονομιάς μέσω της δημιουργίας ψηφιακών αναπαραστάσεων υψηλής ευκρίνειας. Με αυτόν τον τρόπο, εύθραυστα μνημεία και έργα προστατεύονται από άμεση φθορά, ενώ ταυτόχρονα καθίστανται προσβάσιμα σε ευρύτερο κοινό χωρίς να απαιτείται φυσική επαφή με τα πρωτότυπα. Ενδεικτικό παράδειγμα αποτελεί η ψηφιοποίηση των σπηλαιογραφιών στο σπήλαιο Chauvet στη Γαλλία (ψηφιακή περιήγηση[2] και τρισδιάστατη αναπαράσταση[3]), που

[2] archeologie.culture.gouv.fr/chauvet/en/explore-cave?pano=5813
[3] artsandculture.google.com/pocketgallery/QAWBY6bEoEjpLg

επιτρέπει τη μελέτη και προβολή τους χωρίς κίνδυνο αλλοίωσης. Έτσι διασφαλίζεται η μετάδοση αξιόπιστων απεικονίσεων στις επόμενες γενιές.

Αποκατάσταση:

Οι πρόσφατες αναβαθμίσεις στην ΤΝ παρέχουν προηγμένα εργαλεία για την αποκατάσταση έργων πολιτιστικής κληρονομιάς, επιτρέποντας την αναδημιουργία (μέσω εκτίμησης ή πρόβλεψης) και ανακατασκευή ελλειπόντων ή φθαρμένων τμημάτων, καθώς και την ανάδειξη ξεθωριασμένων χρωμάτων με υψηλή ακρίβεια. Χαρακτηριστικά παραδείγματα περιλαμβάνουν την ανακατασκευή των χαμένων τμημάτων[4] του πίνακα The Night Watch (1642) του Rembrandt, καθώς και την αυτόματη και αποτελεσματική ανακατασκευή αρχαιολογικών αντικειμένων που δύσκολο να ανακατασκευαστούν με παραδοσιακές μεθόδους, όπως στο ερευνητικό πρόγραμμα RePAIR[5]. Οι τεχνικές αυτές διευκολύνουν τόσο την ψηφιακή αποκατάσταση όσο και τη βαθύτερη κατανόηση των πολιτιστικών αγαθών.

Στην συνέχεια, επικεντρωνόμαστε στις σύγχρονες εφαρμογές ψηφιακής επεξεργασίας εικόνας σε διαχείριση μωσαϊκών μνημείων. Τα μωσαϊκά αποτελούνται από πολυάριθμες μικρές ψηφίδες, οι οποίες τοποθετούνται με τέτοιο τρόπο ώστε το συνολικό αποτέλεσμα να οδηγεί στον σχηματισμό σύνθετων μοτίβων. Η συγκεκριμένη μορφή τέχνης διαθέτει μακρά ιστορική πορεία και παραμένει σημαντική για την εξέλιξη της σύγχρονης καλλιτεχνικής δημιουργίας. Λόγω της ευαισθησίας των υλικών και της ιδιαίτερης δομικής τους σύνθεσης, τα μωσαϊκά είναι ιδιαίτερα επιρρεπή στη φθορά του χρόνου και συχνά υφίστανται σοβαρές ζημιές, καθιστώντας την ψηφιοποίησή τους αναγκαία για τη διατήρησή τους για τις μελλοντικές γενιές.

Οι παραδοσιακές μέθοδοι ψηφιοποίησης των ψηφίδων, αν και προσφέρουν υψηλή ακρίβεια, είναι ιδιαίτερα χρονοβόρες και απαιτούν εξειδικευμένη γνώση, γεγονός που τις καθιστά δύσκολες στην εφαρμογή τους σε μεγάλης κλίμακας έργα [29,31].

Παραδείγματα:

- **Χειροκίνητη καταγραφή:** Αποτύπωση κάθε ψηφίδας ξεχωριστά με τη βοήθεια χάρακα, μολυβιού ή πινέλου.

- **Σχεδιασμός σε χαρτί:** Σκίτσα και αναπαραστάσεις του μωσαϊκού σε κλίμακα, συχνά με ακριβείς διαστάσεις και χρώματα.

- **Φωτογραφική αποτύπωση και χειροκίνητη αναφορά:** Εκτύπωση φωτογραφιών και σημειώσεις πάνω τους για την καταγραφή των ψηφίδων.

Αντίστοιχα, οι σύγχρονες τεχνικές επιτρέπουν τη δημιουργία ψηφιακών αναπαραστάσεων υψηλής ανάλυσης, παραμένουν όμως σε σημαντικό βαθμό εξαρτημένες από την ανθρώπινη παρέμβαση και περιορίζονται από τη χαμηλή αυτοματοποίησή τους, γεγονός που δυσχεραίνει την εκτεταμένη εφαρμογή τους. Ενδεικτικά:

- **Φωτογραμμετρία:** Δημιουργία τρισδιάστατων ψηφιακών μοντέλων από φωτογραφίες υψηλής ανάλυσης.

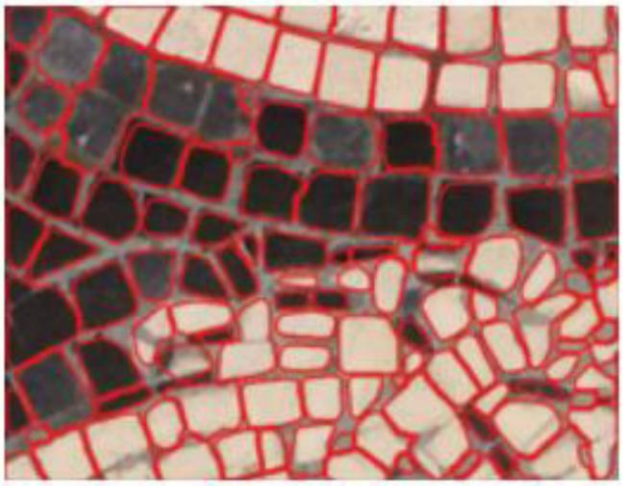
[4] theguardian.com/artanddesign/2021/jun/23/ai-helps-return-rembrandts-the-night-watch-to-original-size
[5] repairproject.eu/

- **3D σάρωση (Laser scanning ή structured light scanning):** Λήψη ακριβών τρισδιάστατων δεδομένων για κάθε ψηφίδα ή ολόκληρο μωσαϊκό.

- **Υπερφασματική απεικόνιση (hyperspectral imaging):** Ανίχνευση υλικών και χρωμάτων με τεχνικές που υπερβαίνουν το ορατό φάσμα φωτός.

Σε αυτό το πλαίσιο, η αυτοματοποιημένη αναγνώριση και κατηγοριοποίηση των ψηφίδων έρχεται να καλύψει το κενό, καθώς επιτρέπει την ακριβή αποτύπωση της δομής των μωσαϊκών, διευκολύνοντας την ψηφιακή αναπαράσταση, την ανάλυση φθορών και την υποβοήθηση της συντήρησης και αποκατάστασης.

Η μελέτη [32] προτείνει μια αυτόματη προσέγγιση βασισμένη στη βαθιά μάθηση, με χρήση αρχιτεκτονικής U-Net, για την τμηματοποίηση των ψηφίδων του δαπέδου μωσαϊκού της εκκλησίας του Αγίου Στεφάνου στο Umm ar Rasas, Ιορδανία, 2021.

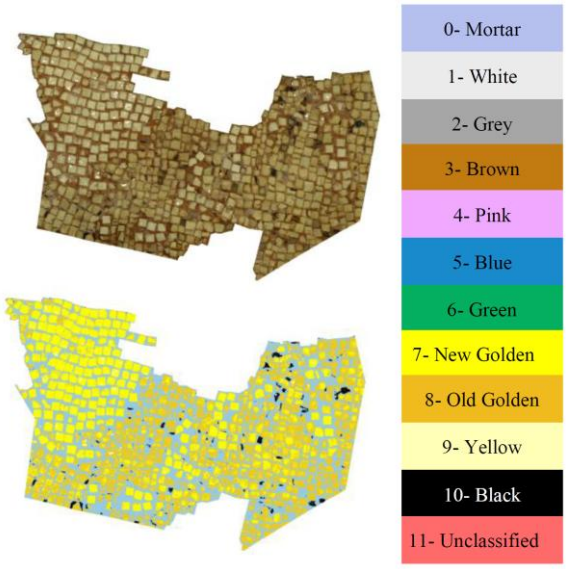


Σχήμα 3.21: Αποτελέσματα προτεινόμενου αλγορίθμου τμηματοποίησης ψηφίδων στην [32]

Η έρευνα [33] αξιοποίησε το εργαλείο τεχνητής νοημοσύνης Segment Anything της META για την αυτόματη τμηματοποίηση των ψηφίδων. Για μεγαλύτερη ακρίβεια, η εφαρμογή περιορίστηκε σε μικρότερες, τοπικές περιοχές της εικόνας, επιτρέποντας τη δημιουργία πιο ακριβών μασκών. Ωστόσο, αυτή η προσέγγιση απαιτεί περεταίρω επεξεργασία. Προβλήματα όπως επικαλύψεις και διπλές μάσκες αντιμετωπίζονται μέσω στατιστικής ανάλυσης των μεγεθών των ψηφίδων και ενός αλγορίθμου επιλογής μάσκας, διασφαλίζοντας την ακρίβεια της τμηματοποίηση.

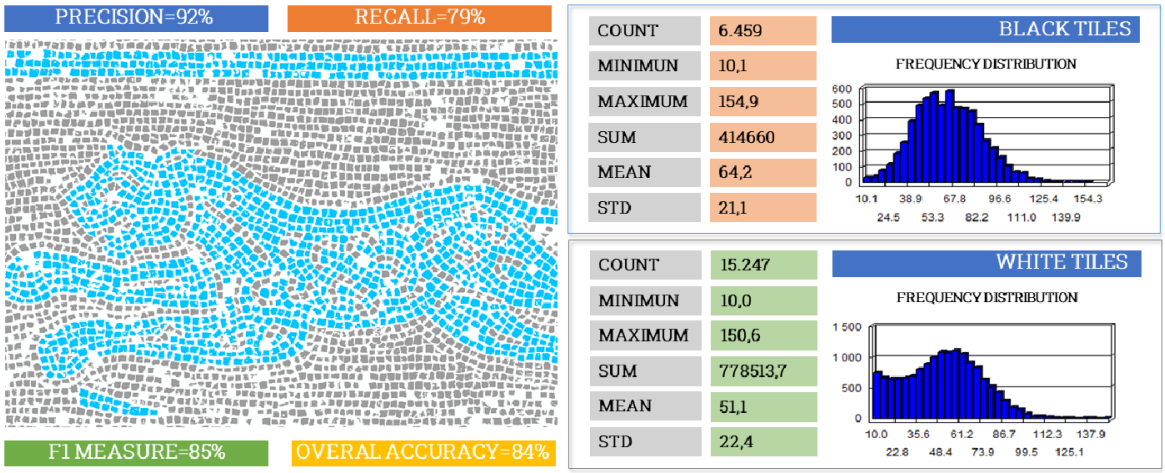
Σχήμα 3.22: Αυτόματη τμηματοποίηση των ψηφίδων ενός μωσαϊκού από τον ναό Saint-Romain-en-Gal, Γαλλία [33]

Το Πανεπιστήμιο Liège στο Βέλγιο έχει επικεντρωθεί στο μωσαϊκό [34-36] που βρίσκεται στο παρεκκλήσι Oratoire carolingien του Germigny-des-Prés, Γαλλία. Η έρευνα [36] παρουσιάζει το Tesserae3D, ένα σύνολο δεδομένων με τρισδιάστατα νέφη σημείων (3D point clouds), ειδικά σχεδιασμένο για την εκπαίδευση και αξιολόγηση μοντέλων μηχανικής μάθησης στην σημασιολογική τμηματοποίηση ψηφίδων μωσαϊκών. Πρόκειται για ένα δημόσια διαθέσιμο σύνολο δεδομένων πολύ υψηλής πυκνότητας (περίπου 502 εκατομμύρια σημεία), το οποίο περιλαμβάνει χρωματική πληροφορία και 11 σημασιολογικές κατηγορίες που καλύπτουν διαφορετικούς τύπους ψηφίδων. Στόχος τους είναι, με βάση ένα σύνολο σημείων που περιγράφουν τη γεωμετρία και το χρώμα κάθε ψηφίδας, το σύστημα να αποδίδει αυτόματα μία ετικέτα σε κάθε σημείο, 2021.



Σχήμα 3.23: Πάνω: Αρχικό νέφος σημείων και κάτω: Η ταξινόμηση των σημείων με βάση το χρώμα (1-11) με χρήση του αλγορίθμου Random Forest [36]

Η έρευνα [37] περιγράφει μια μεθοδολογία για την αυτόματη ταξινόμηση των ψηφίδων ενός μονοχρωματικού ρωμαϊκού μωσαϊκού από τον αρχαιολογικό χώρο της Saltara, χρησιμοποιώντας εργαλεία και αλγόριθμους που υλοποιούνται σε λογισμικό γεωγραφικού συστήματος πληροφοριών (ArcGIS - Geographic Information System), 2024.



Σχήμα 3.24: Αυτόματης ανακατασκευής στο περιβάλλον ArcGIS και υπολογισμός των επιδόσεων και στατιστικών δεικτών που προέκυψαν [37]

Ψηφιακή Ανακατασκευή:

Η έρευνα [38] εξετάζει τη χρήση τεχνητής νοημοσύνης, με το εργαλείο DALL-E2 της OpenAI, για την ανακατασκευή αρχαίων μωσαϊκών με ελλείποντα τμήματα, δείχνοντας ότι η ΤΝ μπορεί να αναγνωρίσει βασικά χαρακτηριστικά και να δημιουργήσει αληθοφανής αναπαραστάσεις των σκηνών.

Σχήμα 3.25: Ψηφιακή ανακατασκευή μωσαϊκών: Αριστερά, το «Lion attacking an Onager», που βρέθηκε στην Τυνησία, και δεξιά, το ψηφιδωτό του Μεγάλου Αλεξάνδρου, ανακαλύφθηκε στην έπαυλη του Φαύνου στην Πομπηία [38]

Αναγνώριση Μοτίβων:
Η έρευνα [39] προτείνει μια τεχνική βασισμένη στη βαθιά μάθηση και, πιο συγκεκριμένα, σε CNNs, η οποία εντοπίζει και ταξινομεί τα γεωμετρικά σχήματα σε μωσαϊκά.

Σχήμα 3. 26: Αριστερά: Ρωμαϊκό μωσαϊκό που ανακαλύφθηκε στο Savignano sul Panaro και Δεξιά: Απεικόνιση των γεωμετρικών σχημάτων του μωσαϊκού [39]

Τρισδιάστατη (3D) Αναπαράσταση:
Ένα παράδειγμα αποκατάστασης με ανθρώπινη παρέμβαση, και ψηφιακής αναγνώριση χρωμάτων και 3D αναπαράστασης του μωσαϊκού με τη σκηνή της αρπαγής της Ευρώπης, το οποίο εκτίθεται στο Universalmuseum Joanneum στο Graz [29].

Σχήμα 3.27: Αριστερά: Μωσαϊκό με τη σκηνή της αρπαγής της Ευρώπης, το οποίο εκτίθεται στο Universalmuseum Joanneum στο Graz. Δεξιά: Στιγμιότυπο από το βίντεο της τρισδιάστατης απεικόνισης (επάνω) και τρισδιάστατη διαδικασία απεικονισμένη με σωματίδια (κάτω) [29]

Μετά την παρουσίαση πρόσφατων μελετών, αξίζει να σημειωθούν ορισμένες προκλήσεις που αντιμετωπίζει η εφαρμογή της τεχνητής νοημοσύνης σε αυτόν τον τομέα [30]:

- **Εκπαίδευση μοντέλων ΒΜ:** απαιτεί μεγάλο όγκο δεδομένων, που στην πράξη είναι συχνά περιορισμένα. Ακόμη και αν συνδυαστούν παρόμοια δεδομένα από διαφορετικές πηγές, υπάρχει υψηλή πιθανότητα να εμφανίζουν διαφοροποιήσεις που μπορούν να επηρεάσουν την ακρίβεια και την αξιοπιστία των μοντέλων.

- **Διατήρηση αυθεντικότητας των έργων:** αποτελεί κρίσιμη προτεραιότητα, καθώς οι προτεινόμενες επεμβάσεις ενδέχεται να αποκλίνουν από την ιστορική πραγματικότητα.

- **Προστατευόμενα δεδομένα:** δεδομένα που προέρχονται από πολιτιστικά αγαθά συχνά προστατεύονται από τη νομοθεσία, γεγονός που απαιτεί η χρήση της τεχνητής νοημοσύνης να συμμορφώνεται πλήρως με τα ισχύοντα νομικά πλαίσια, προλαμβάνοντας κακόβουλη ή λανθασμένη αναπαραγωγή τους.

3.1.5 ΒΙΒΛΙΟΓΡΑΦΙΑ

1. J. Jing, S. Liu, G. Wang, W. Zhang, C. Sun, “Recent advances on image edge detection: A comprehensive review”, 2022, Neurocomputing, Vol. 503, pp. 259–271, 2022, doi: 10.1016/j.neucom.2022.06.083.
2. Richard Szeliski, “Computer Vision Algorithms and Application Second Edition”, 2022, Springer, doi: 10.1007/978-3-030-34372-9
3. Sun R, Lei T, Chen Q, Wang Z, Du X, Zhao W and Nandi AK “Survey of Image Edge Detection”, 2022, Frontiers in Signal Processing 2:826967, doi: 10.3389/frsip.2022.826967
4. Jain, R., Kasturi, R., Schunck, B., “Machine Vision”, McGraw-Hill Education 1995, ISBN 0-07-032018-7
5. R. C. Gonzalez & R. E. Woods, “Digital Image Processing Fourth Edition”, Pearson Education Limited, ISBN 978-0-13-335672-4
6. Mahalle, G. A. & Shah, A. M. “FPGA Implementation of Gradient Based Edge Detection Algorithms”, 2017, IJIRCCE, doi: 10.15680/IJIRCCE.2017. 0505087
7. Balochian, S., Baloochian, H., “Edge detection on noisy images using Prewitt operator and fractional order differentiation”, 2022, Multimedia Tools and Applications, vol. 81, pp. 1–12, doi: 10.1007/s11042-022-12011-1
8. Elharrouss, O., Al-Maadeed, S., Subramanian, N., Ottakath, N., Almaadeed, N., Himeur, Y., “Panoptic Segmentation: A Review”, 2021, arXiv:2111.10250
9. Τομαρά Ηλίας, “Ανίχνευση μορφολογικών χαρακτηριστικών βλαστοκύστεων με χρήση μοντέλων σημασιολογικής τμηματοποίησης”, 2025, Μεταπτυχιακή Εργασία, Πανεπιστήμιο Πειραιώς.
10. Xu, Y., Quan, R., Xu,W., Huang, Y., Chen, X., Liu, F. “Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches”, 2024, Bioengineering, 11, 1034, doi: 10.3390/bioengineering11101034
11. Brar, K. K., Goyal, B., Dogra, A., Mustafa, M. A., Majumdar, R., Alkhayyat, A., Kukreja, V., “Image segmentation review: Theoretical background and recent advances”, 2025, Information Fusion, vol. 114, p. 102608, 2025, doi: 10.1016/j.inffus.2024.102608
12. Yu, Y.,Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., Yang, T., Gao, M. “Techniques and Challenges of Image Segmentation: A Review”, Electronics 2023, 12, 1199, doi: 10.3390/electronics12051199
13. Bishop M. C. & Bishop H., “Deep Learning Foundations and Concepts”, 2024, Springer, doi: 10.1007/978-3-031-45468-4
14. Taye, M.M., “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions”, 2023, Computers, 12, 91, doi: 10.3390/computers12050091
15. Xu P, Wang J, Jiang Y and Gong X, “Applications of artificial intelligence and machine learning in image processing”, 2024, Frontiers in Materials, 11:1431179, doi: 10.3389/fmats.2024.1431179
16. S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos, “Image Segmentation Using Deep Learning: A Survey”, 2022, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, No. 7, pp. 3523–3542, doi: 10.1109/TPAMI.2021.3059968
17. Matthew, B., Genc, B., Huang, Y., Belter, D., Tang, Z., “Comparative Analysis of Popular Semantic Segmentation Architectures: U-Net, DeepLab, and FCN”, 2024
18. Krizhevsky, A., Sutskever, I., Hinton, G.E., “ImageNet Classification with Deep Convolutional Neural Networks”, 2012, Advances in Neural Information Processing Systems, Vol. 25, 2012, doi: 10.1145/3065386
19. Simonyan K. and Zisserman A., “Very Deep Convolutional Networks for Large-Scale Image Recognition”, 2015, arXiv:1409.1556
20. Ronneberger, O., Fischer, P., Brox, T. “U-Net: Convolutional Networks for Biomedical Image Segmentation”, 2015, arXiv:1505.04597
21. Li, Z., Zhang, H., Li, Z., and Ren, Z., “Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation”, 2022, Appl. Sci., 12, 7149, doi: 10.3390/app12147149
22. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., ”Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs”, 2016, arXiv:1412.7062
23. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”, 2018, arXiv:1802.02611.

24. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, 2021, arXiv:2010.11929.
25. Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P., Wang, L., “Vision Transformers for Image Classification: A Comparative Survey”, 2025, Technologies, 13, 32, doi: 10.3390/technologies13010032
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., “Attention Is All You Need”, 2017, arXiv:1706.0376
27. Gao, Y., Liang, J., Yang, J., “Color Palette Generation From Digital Images: A Review”, 2025, Color Research & Application, vol. 50, no. 3, doi: 10.1002/col.2297
28. Chang, Y., Mukai, N., “Color Feature Based Dominant Color Extraction”, 2022, IEEE Access, vol. 10, pp. 93055–93061, 2022, doi: 10.1109/ACCESS.2022.3202632
29. Oštir, G., Javoršek, D., Stergar, P., Koˇcevar, T.N., Nestorovi´c, A., Gabrijelˇciˇc Tomc, H., “Graphic Reconstruction of a Roman Mosaic with Scenes of the Abduction of Europa”, 2024, Appl. Sci., 14, 3931, doi: 10.3390/app14093931
30. Δερμεντζόπουλος Α. & Σουλιώτου Α. Ζ., “Η συμβολή της Τεχνητής Νοημοσύνης στη διαφύλαξη και στην επαύξηση της εμπειρίας της Πολιτιστικής Κληρονομιάς”, 2025, Journal of Culture in Tourism, Art and Education, Vol. 5, Issue 1, doi: 10.26220/cul.5323
31. Doria, E. and Picchio, F., “TECHNIQUES FOR MOSAICS DOCUMENTATION THROUGH PHOTOGRAMMETRY DATA ACQUISITION. THE BYZANTINE MOSAICS OF THE NATIVITY CHURCH”, 2020, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., 965–972, doi: 10.5194/isprs-annals-V-2-2020-965-2020, 2020.
32. Felicetti, A., Paolanti, M., Zingaretti, P., Pierdicca, R., & Malinverni, E. S., “Mo.Se.: Mosaic image segmentation based on deep cascading learning. Virtual Archaeology Review”, 2021, 12(24), 25–38, doi: 10.4995/var.2021.14179.
33. Jean Mélou, Antoine Laurent, Nikola Dabic, Romain Poirier, Ani Danielyan, et al.., “Deep-learning based Detection and Segmentation in Archaeology”, 2025, Digital Horizons: Embracing heritage in an evolving world, CAA, May 2025, Athènes, Greece. hal-05058754.
34. Poux, F.; Neuville, R.; Billen, R., “Point Cloud Classification of Tesserae from Terrestrial Laser Data Combined with Dense Image Matching for Archaeological Information Extraction”, 2017, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-2/W2, pp. 203–211, doi: 10.5194/isprs-annals-IV-2-W2-203-2017.
35. Poux, F., Neuville, R., Van Wersch, L., Nys, G.-A., & Billen, R., “3D Point Clouds in Archaeology: Advances in Acquisition, Processing and Knowledge Integration Applied to Quasi-Planar Objects”, 2017. Geosciences, 7(4), 96. doi: 10.3390/geosciences7040096.
36. Kharroubi, A.; Van Wersch, L.; Billen, R.; Poux, F., “TESSERAE3D: A Benchmark for Tesserae Semantic Segmentation in 3D Point Clouds”, 2021, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. V-2-2021, pp. 121–128, doi: 10.5194/isprs-annals-V-2-2021-121-2021.
37. Alfio, V.; Pepe, M.; Costantino, D.; Baratin, L., “GIS Approach for Mosaic Tesserae Recognition by Supervised Image Classification”, 2024, Studies in Digital Heritage, Vol. 8, pp. 1–14, doi: 10.14434/sdh.v8i1.36724
38. Moral-Andrés, F.; Merino Gómez, E.; Reviriego, P.; Lombardi, F., “Can Artificial Intelligence Reconstruct Ancient Mosaics?”, 2022, doi: 10.48550/arXiv.2210.06145.
39. Ghosh, M., Obaidullah, S.M., Gherardini, F., Zdimalova, M.“Classification of Geometric Forms in Mosaics Using Deep Neural Network”, 2021, J. Imaging, 7, 149, doi: 10.3390/jimaging7080149