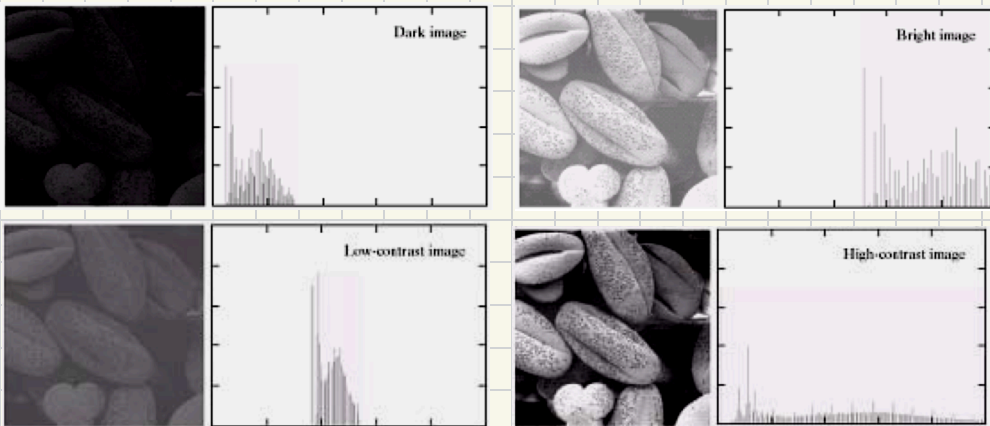


21. Περιγράψτε σύντομα για κάθε μία από τις ανωτέρω εικόνες πώς νομίζετε ότι θα είναι το ιστόγραμμα της. Δικαιολογήστε την απάντησή σας.



23. Θέλετε να εκπαιδεύσετε δύο βαθιά συνελκτικά δίκτυα να κατηγοριοποιούν, το πρώτο εικόνες από διάφορα κατοικίδια ζώα, το δε δεύτερο ιατρικές εικόνες ακτινογραφιών σε δύο κατηγορίες: ασθενών και υγιών ατόμων. Διαθέτετε λίγες εικόνες τόσο στο πρώτο όσο και στο δεύτερο πρόβλημα. Βρίσκετε στο διαδίκτυο το βαθύ συνελκτικό δίκτυο που έχει εκπαιδευτεί στο CIFAR-100. Πώς θα το χρησιμοποιούσατε σε κάθε μία από τις δύο περιπτώσεις;

Το δίκτυο που έχει εκπαιδευτεί στο CIFAR-100 περιέχει γνώση για γενικά αντικείμενα και εικόνες φυσικού κόσμου. Η αξιοποίησή του εξαρτάται από το πόσο σχετικό είναι το πρόβλημα:

(α) Κατοικίδια:

Οι εικόνες κατοικίδιων είναι παρόμοιες με τις εικόνες του CIFAR-100 (ζώα, φυσικές συνθήκες, RGB). Άρα, μπορούμε να χρησιμοποιήσουμε fine-tuning:

- Κρατάμε τα πρώτα επίπεδα (που μαθαίνουν γενικά χαρακτηριστικά όπως ακμές και υφές).
- Επανεκπαιδεύουμε τα τελευταία επίπεδα (που είναι υπεύθυνα για ταξινόμηση) πάνω στις δικές μας κατηγορίες. Αυτό είναι αποτελεσματικό ακόμα και με μικρό αριθμό εικόνων.

(β) Ακτινογραφίες:

Οι ιατρικές εικόνες (ακτινογραφίες) είναι διαφορετικές: grayscale, διαφορετική υφή, διαφορετικά χαρακτηριστικά.

Σε αυτή την περίπτωση, δεν είναι ιδανικό να κάνουμε fine-tuning σε όλο το δίκτυο, γιατί τα αρχικά φίλτρα μπορεί να μην είναι χρήσιμα.

→ Προτιμούμε να το χρησιμοποιήσουμε ως feature extractor:

Παγώνουμε όλα τα επίπεδα (δεν τα επανεκπαιδεύουμε).

Παίρνουμε τις εξόδους ενός ενδιάμεσου ή τελικού layer και εκπαιδεύουμε μόνο έναν απλό ταξινομητή (π.χ. logistic regression ή fully connected layer). Αυτό μειώνει τον κίνδυνο υπερπροσαρμογής.

21. Χρησιμοποιήσατε μια χαλασμένη κάμερα για να πάρετε μια φωτογραφία, με αποτέλεσμα, αν και η φωτογραφία ήταν αρκετά φωτεινή - χωρίς κανένα μαύρο εικονοστοιχείο - η εικόνα που πήρατε (Εικόνα 1) έχει μαύρα, με τυχαίο τρόπο, το 10% των εικονοστοιχείων της. Εφαρμόζετε ένα φίλτρο ενδιάμεσης τιμής (1 γραμμή × 3 στήλες) στην εικόνα αυτή, με αποτέλεσμα να βελτιώσετε την ποιότητά της (Εικόνα 2). Δείξτε ότι σε αυτή την εικόνα το 2.8% των εικονοστοιχείων είναι μαύρα.



διακριτή τιμ: $z_i = \begin{cases} 0, \text{μαύρο pixel} \\ 1, \text{άσπρο pixel} \end{cases}$ } αρχική εικόνα

$P(z=0) = 0.1$ και $P(z=1) = 0.9$

Για να είναι μαύρο ένα pixel στην τελική εικόνα πρέπει να έχουμε $\{MM, AM, MA, MM\}$

$P(MM) = 0.1^2 = 0.001$

$P(AM) = P(MA) = P(MM) = 0.1^2 \cdot 0.9 = 0.009$

$P(\text{μαύρο}) = 0.001 + 3 \cdot 0.009 = 0.028 = 2.8\%$

Κανονική 2012 - Θ3

Σας δίδονται οι 5 εικόνες (α-ε) ενός ναού που έχουν προέλθει από εφαρμογή κάποιου φίλτρου στην αρχική εικόνα. Επίσης σας δίδονται τα 5 φάσματα των φίλτρων που έχουν δημιουργήσει τις εικόνες, σε τυχαία μορφή. Αντιστοιχίστε τα φίλτρα με τις εικόνες που δημιούργησαν, δίνοντας μια σύντομη εξήγηση.

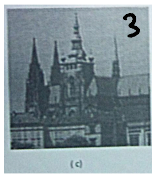
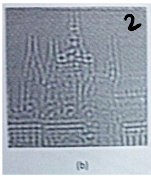
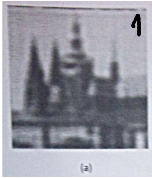
θόλωμα

θορύβοι

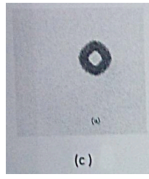
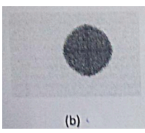
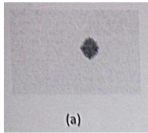
θόλωμα

περίγραμμα

θορύβοι



1. e
2. a
3. d
4. c
5. b



a, b : high-pass
κρατάει παραπάνω πληροφορίες

c: band-pass

d, e: low-pass
κρατάει ανακ

πιο θωλή

Κανονική 2013 - Θ2

Ταιριάζετε τους διάφορους τελεστές που αναφέρονται παρακάτω με το αποτέλεσμα κάθε βέλους στο επόμενο σετ από εικόνες.



- A. Φίλτρο μέγιστης τιμής
- B. Εξίσωση ιστογράμματος
- Γ. Λαπλασιανός τελεστής
- Δ. Φίλτρο μέσου όρου
- Ε. Φίλτρο ενδιάμεσης τιμής
- ΣΤ. Ανιχνευτής ακμών Sobel (1 κατεύθυνση)
- Ζ. Φίλτρο Όξυνσης
- Η. Μετασχηματισμός Φωτεινότητας

1. A
2.
3.
4. E
5.
6.
7. ΣΤ
8. Η

1. Αν έχω αυτό το τάδε φίλτρο τι περιμένετε να γίνει αν εφαρμόσουμε στην εικόνα;

- **Low-pass** (όλα τα στοιχεία θετικά, άθροισμα = 1): θολώνει (smoothing), μειώνονται οι αιχμηρές ακμές.
- **High-pass** (θετικό κεντρικό βάρος, αρνητικά πέριξ): αναδεικνύει ακμές (edge detection/sharpening).
- **Κατευθυντικό (Sobel/Prewitt)**: τονίζει ακμές μόνο σε μια διεύθυνση (οριζόντιες ή κάθετες).
- **Median filter**: αφαιρεί θόρυβο αλατοπιπίτερου χωρίς να θολώνει τις ακμές.

2. Τι μορφή έχουν οι εικόνες (χαρακτηριστικά). Τι είναι η έγχρωμη εικόνα;

- Η **ψηφιακή** εικόνα είναι ουσιαστικά ένας διδιάστατος πίνακας από εικονοστοιχεία (pixels), όπου κάθε pixel έχει μια τιμή έντασης φωτεινότητας (σε κλίμακα γκρι) ή ένα σύνολο τιμών (σε έγχρωμη). Σε ασπρόμαυρη (grayscale) εικόνα, κάθε pixel αποθηκεύεται με έναν αριθμό (π.χ. 0–255 για 8-bit) που αντιστοιχεί στην αντίστοιχη απόχρωσή του.
- Η **έγχρωμη** εικόνα (RGB) αποτελείται από τρία κανάλια (Red, Green, Blue). Κάθε pixel περιέχει τριάδα τιμών (π.χ. $[R, G, B] = [0-255, 0-255, 0-255]$), η οποία προσδιορίζει το χρώμα του.

3. Ιδιότητες ανθρώπινης όρασης να εκμεταλλευτούμε για να αναλύσουμε το περιεχόμενο και κωδικοποίηση εικόνας

- **Ανάλυση συχνότητων**: Το ανθρώπινο μάτι είναι πιο ευαίσθητο στις χαμηλές συχνότητες (μεγάλα, αργά σχήματα) παρά στις υψηλές (λεπτές λεπτομέρειες). Άρα στην κωδικοποίηση διατηρούμε πρώτα τις χαμηλές συχνότητες και συμπεριέχουμε περισσότερο τις υψηλές.
- **Χρωματική αντίληψη**: Οι άνθρωποι αντιλαμβάνονται πιο καλά διαφορές στη φωτεινότητα παρά στο χρώμα. Γι' αυτό στις έγχρωμες εικόνες (π.χ. σε YCbCr) μπορεί να μειωθεί η ανάλυση (subsampling) στα κανάλια χρώματος χωρίς μεγάλη οπτική απώλεια.
- **Πλεονάζουσα πληροφορία (masking)**: Σε περιοχές με πολύπλοκη υφή ή έντονη φωτεινότητα, μικρές παραμορφώσεις δεν γίνονται εύκολα αντιληπτές. Οι κωδικοποιητές εκμεταλλεύονται το φαινόμενο, κάνοντας μεγαλύτερη συμπίεση στις "θορυβώδεις" περιοχές.
- **Έμφαση στις ακμές**: Οι αιχμηρές ακμές και οι μεταβάσεις χρώματος παίζουν κρίσιμο ρόλο στην αναγνώριση αντικειμένων. Στον έλεγχο ποιότητας, διατηρούνται περισσότερο τα δεδομένα γύρω από τις ακμές (π.χ. με φίλτρα υψηλών συχνότητων).
- **Φωτομετρική αντίληψη (contrast sensitivity)**: Το μάτι αντιλαμβάνεται καλύτερα μεταβολές κοντά σε μεσαία επίπεδα φωτεινότητας. Η κωδικοποίηση προσαρμόζει το quantization ώστε να μειώσει θόρυβο κυρίως στις πολύ σκοτεινές ή πολύ φωτεινές περιοχές, όπου η αντίληψη είναι πιο "αμβλεία."
- **Χωρική αλληλεξάρτηση (spatial correlation)**: Γειτονικά pixels έχουν συχνά παρόμοια τιμή. Αντί να κωδικοποιήσουμε ξεχωριστά κάθε pixel, μπορούμε να αναλύσουμε διάφορες χωρικές συχνότητες (π.χ. με DCT ή wavelets) και να συμπίεσουμε την ομοιογένεια σε ομαλές περιοχές.
- **Χρονική ευαισθησία (motion masking)**: Όταν υπάρχει κίνηση, οι μικρές λεπτομέρειες σε κινούμενες περιοχές είναι πιο δύσκολο να διακριθούν. Στο video coding (MPEG) οι διαφορές από καρέ σε καρέ κωδικοποιούνται πιο χονδρικά όπου υπάρχει έντονη κίνηση, καθώς το ανθρώπινο μάτι δεν συλλαμβάνει τόσο εύκολα τις λεπτομέρειες σε γρήγορη κίνηση.

4. Συσχέτιση εικόνας στο πεδίο του χώρου + πεδίο της συχνότητας

χαμηλές συχνότητες → αντιστοιχούν στο background, σε ομαλές περιοχές της εικόνας

υψηλές συχνότητες → αντιστοιχούν στις άκρες και στον θόρυβο

Αυτή είναι η διαφορά της ανάλυσης των εικόνων από άλλα σήματα (π.χ. μονοδιάστατα σήματα ή η φωνή) αφού εδώ μας ενδιαφέρει μόνο το περιεχόμενο στις χαμηλές συχνότητες αλλά και στις ψηλές συχνότητες που είναι οι ακμές

5. Ποια ιδιότητα του JPEG χρησιμοποιούμε στο βίντεο;

Η ανθρώπινη όραση δεν αντιλαμβάνεται αλλαγές που συμβαίνουν πάνω από 50–60 φορές το δευτερόλεπτο. Αυτό επιτρέπει τεχνικές όπως interlacing και χρονική υποδειγματοληψία στο βίντεο, μειώνοντας τα δεδομένα χωρίς ορατή απώλεια ποιότητας.

6. Μετ/μος Fourier, DCT, Wavelet, ημιτονο, τι ακριβώς είναι ο καθένας και που χρησιμοποιείται;

- **FT (Fourier Transform)**: Αναλύει το σήμα σε άπειρες ημιτονικές/συνημιτονικές συναρτήσεις. Δίνει μόνο πληροφορία συχνότητας, χωρίς τοπικότητα. Χρησιμοποιείται σε γενική φασματική ανάλυση, όχι σε JPEG/MPEG.
- **DCT (Διακριτός Μετασχηματισμός Συνημιτόνου)**: Παρόμοιος με τον Fourier αλλά μόνο με συνημίτονα (πραγματικές τιμές). Συγκεντρώνει την ενέργεια στις χαμηλές συχνότητες. Χρησιμοποιείται σε JPEG και MPEG για συμπίεση (σε blocks 8x8).
- **Wavelet Transform**: Αναλύει το σήμα σε κυματάκια (wavelets), με πληροφορία και σε χώρο και σε συχνότητα. Λειτουργεί σε πολλαπλές κλίμακες (multi-scale). Χρησιμοποιείται στο JPEG 2000 — προσφέρει υψηλή συμπίεση χωρίς block artifacts.
- **DST (Discrete Sine Transform)**: Μετασχηματισμός μόνο με ημιτόνια. Λιγότερο διαδεδομένος, χρησιμοποιείται σε ειδικές περιπτώσεις.

7. Βασικές ιδιότητες JPEG & MPEG – Εκτίμηση κίνησης

- Το **JPEG** εφαρμόζεται σε στατικές εικόνες και περιλαμβάνει: DCT → Κβαντισμό → Εντροπική κωδικοποίηση, με στόχο τη συμπίεση διατηρώντας την ποιότητα.
- Το **MPEG** επεκτείνει αυτή τη λογική στο χρόνο, για βίντεο. Χρησιμοποιεί τρία είδη καρέ: I-frames: ανεξάρτητα (σαν JPEG), P-frames: πρόβλεψη από προηγούμενο καρέ, B-frames: πρόβλεψη από προηγούμενο & επόμενο καρέ.
- Κατά την **εκτίμηση κίνησης**, ο πομπός εντοπίζει για κάθε block τη μετατόπιση του (διάνυσμα κίνησης) και στέλνει αυτό μαζί με το σφάλμα πρόβλεψης. Η εκτίμηση γίνεται πάνω στις ανακατασκευασμένες εικόνες, όχι στις αρχικές, ώστε δέκτης και πομπός να παραμένουν πλήρως συγχρονισμένοι χωρίς σωρευτικά λάθη.

8. Ποιος είναι ο ρόλος του κβαντιστή στην κωδικοποίηση και πώς επηρεάζει την ποιότητα και τη συμπίεση του σήματος;

Ο κβαντιστής μειώνει την ακρίβεια των τιμών (π.χ. συντελεστών DCT) στρογγυλοποιώντας τις σε λιγότερα επίπεδα. Έτσι επιτυγχάνεται μεγαλύτερη συμπίεση, αλλά με απώλεια ποιότητας (lossy).

Ο βαθμός κβαντισμού καθορίζει το πόση λεπτομέρεια διατηρείται:

- Ισχυρός κβαντισμός → μικρότερο αρχείο, περισσότερη απώλεια
- Ασθενής κβαντισμός → καλύτερη ποιότητα, μεγαλύτερο μέγεθος

Στο MPEG, εφαρμόζεται τόσο στους συντελεστές DCT όσο και στο σφάλμα πρόβλεψης, κρατώντας κυρίως την πληροφορία με μεγαλύτερη οπτική σημασία.

9. Ποια είναι τα σήματα (πληροφορία) που στέλνει ο κωδικοποιητής MPEG στο δέκτη;

Ο MPEG κωδικοποιητής στέλνει διαφορετικές πληροφορίες ανάλογα με τον τύπο του καρέ:

- I-frames: Στέλνεται ολόκληρη η εικόνα, χωρίς motion vectors (όπως στο JPEG).
- P-frames: Στέλνονται τα motion vectors και το σφάλμα πρόβλεψης (residual).
- B-frames: Στέλνονται μόνο motion vectors (πρόβλεψη από δύο καρέ).

Έτσι επιτυγχάνεται σημαντική χρονική συμπίεση, χωρίς αισθητή απώλεια ποιότητας.

10. Lossless vs Lossy τεχνικές

#	Τεχνική Κωδικοποίησης	Τύπος	Σχόλιο
1	LZW / GIF Coding	Lossless	Συχνά σε GIF & TIFF
2	Difference Encoding / Quantisation	Lossy	Η απώλεια προκύπτει από τον κβαντισμό
3	Discrete Cosine Transform (DCT)	Lossless*	Μόνο αν δεν περιλαμβάνει κβαντισμό
4	Arithmetic Coding	Lossless	Στατιστική κωδικοποίηση
5	Vector Quantisation	Lossy	Χαρακτηριστικά block-based συμπίεσης
6	Huffman Coding	Lossless	Κλασικός lossless αλγόριθμος
7	Pattern Substitution	Lossless	Αντικαθιστά επαναλαμβανόμενα μοτίβα
8	Wavelet Coding	Lossless*	Χωρίς κβαντισμό παραμένει lossless
9	JPEG Coding	Lossy	Περιλαμβάνει DCT + Κβαντισμό + Huffman
10	Run Length Encoding	Lossless	Ιδανική για απλές επαναλήψεις

- **Lossless:** Δεν υπάρχει καμία απώλεια πληροφορίας. Επιτρέπει πλήρη ανακατασκευή της αρχικής εικόνας. Χρησιμοποιείται σε ιατρικές/επιστημονικές εφαρμογές. Βασίζεται σε εντροπική κωδικοποίηση και εκμεταλλεύεται πλεονασμούς (π.χ. όμοια γειτονικά pixels).
- **Lossy:** Επιτρέπει απώλειες πληροφορίας για να επιτύχει μεγαλύτερη συμπίεση. Χρησιμοποιείται σε εικόνες/βίντεο. Εκμεταλλεύεται την ανθρώπινη όραση: το μάτι είναι λιγότερο ευαίσθητο σε λεπτές χρωμικές ή χρωματικές μεταβολές, άρα η απώλεια δεν γίνεται ορατή.

11. Τι επιδιώκουμε στην ανάλυση εικόνας;

Στόχος της ανάλυσης εικόνας είναι η εξαγωγή πληροφορίας από το περιεχόμενο. Κύρια εφαρμογή: ανίχνευση ακμών, δηλαδή εντοπισμός απότομων αλλαγών φωτεινότητας. Οι ακμές καθορίζουν τα όρια αντικειμένων, από τα οποία μπορούμε να αναγνωρίσουμε σχήματα, να εξάγουμε χαρακτηριστικά και να προχωρήσουμε σε ταξινόμηση, αναγνώριση ή παρακολούθηση αντικειμένων.

12. Τελεστές ανίχνευσης ακμών – Στόχος, πλεονεκτήματα, σύγκριση

Οι τελεστές ακμών εντοπίζουν περιοχές με απότομη μεταβολή φωτεινότητας. Οι βασικότεροι είναι:

- **Prewitt / Sobel:** Βασίζονται σε παραγώγους πρώτης τάξης. Υπολογίζουν την κλίση φωτεινότητας κατά x και y. Είναι απλοί και γρήγοροι, αλλά ευαίσθητοι σε θόρυβο.
- **Canny:** Πιο σύνθετος τελεστής. Χρησιμοποιεί Gaussian εξομάλυνση + 1η παράγωγος + μη μέγιστη καταστολή + κατώφλια. Παρέχει λεπτή και ακριβή ανίχνευση με λίγα ψευδώς θετικά.
- **Gaussian:** Δεν ανιχνεύει ακμές, αλλά χρησιμοποιείται για αποθορυβοποίηση πριν την ανίχνευση.
- **Laplacian (2ης τάξης):** Εντοπίζει ακμές στα zero-crossings της δεύτερης παραγώγου. Πιο ευαίσθητος στον θόρυβο, συνήθως συνδυάζεται με Gaussian (Log).

Σύγκριση: 1ης τάξης: βρίσκουν μέγιστα στην κλίση (Prewitt, Sobel, Canny) / 2ης τάξης: βρίσκουν σημεία αλλαγής καμπυλότητας (Laplacian)

Επιλογή τελεστή: Sobel/Prewitt → απλότητα, Canny → ακρίβεια, Laplacian → καμπυλότητα, σε συνδυασμό με εξομάλυνση

13. Ιεραρχική αναπαράσταση τελεστών – Ανάλυση σε πολλαπλά scales

Η **ιεραρχική αναπαράσταση** βασίζεται στην ιδέα ότι τα χαρακτηριστικά της εικόνας πρέπει να εντοπίζονται σε διαφορετικές κλίμακες (scales), από λεπτομέρειες έως πιο γενικά σχήματα. Αυτό επιτυγχάνεται με **εφαρμογή Gaussian φίλτρων** με αυξανόμενη τυπική απόκλιση (σμίκρυνση/θόλωμα) και χρήση τελεστών (π.χ. Sobel, Laplacian) σε κάθε επίπεδο. Έτσι προκύπτει μια πυραμίδα εικόνων, όπου κάθε επίπεδο αναδεικνύει διαφορετικού μεγέθους χαρακτηριστικά. Η προσέγγιση αυτή στηρίζεται στο **scale-space** και στους **wavelet μετασχηματισμούς**, επιτρέποντας ανάλυση τόσο σε χώρο όσο και σε συχνότητα, με στόχο ανίχνευση ανεξαρτήτως μεγέθους.

14. Βαθιά νευρωνικά δίκτυα – Βασική διαφορά

Στα βαθιά νευρωνικά δίκτυα (deep neural networks), τα χαρακτηριστικά εξάγονται αυτόματα μέσα από τα διαδοχικά επίπεδα του δικτύου, χωρίς να ορίζονται χειροκίνητα. Κάθε επίπεδο μαθαίνει πιο αφηρημένες και σύνθετες αναπαραστάσεις, ξεκινώντας από απλά μοτίβα (π.χ. ακμές) και φτάνοντας σε υψηλού επιπέδου έννοιες (π.χ. μορφές ή αντικείμενα). Η αυτόματη εξαγωγή χαρακτηριστικών είναι το κύριο πλεονέκτημα έναντι των παραδοσιακών μεθόδων.

15. Convolutional Neural Networks (CNN) – Αρχιτεκτονική

Σε ένα CNN, κάθε φίλτρο εφαρμόζεται σε τοπικές περιοχές της εισόδου και παράγει ένα feature map (ή activation map). Όσα περισσότερα φίλτρα, τόσο μεγαλύτερο το βάθος της εξόδου. Το receptive field ενός νευρώνα είναι το κομμάτι της εικόνας που «βλέπει». Με το βάθος, το receptive field μεγαλώνει. Οι διαστάσεις του feature map εξαρτώνται από: Μέγεθος εισόδου, Μέγεθος φίλτρου, Stride, Padding. Αν έχουμε π.χ. 64 φίλτρα, προκύπτουν 64 feature maps → έξοδος με βάθος 64. Ο αριθμός παραμέτρων για κάθε φίλτρο: $\text{filter size} = w \times h \times \text{depth} + 1$ (bias). Συνολικός παράμετροι = αριθμός φίλτρων $\times$ (μέγεθος φίλτρου + bias).

16. Ιδιότητες υπερπερατού φίλτρου

Αν ένα φίλτρο έχει συντελεστές με διαφορετικά πρόσημα που αθροίζονται στο μηδέν, τότε είναι υπερπερατό και χρησιμοποιείται για όξυνση εικόνων και ανίχνευση ακμών.

- Σε ομοιόμορφες περιοχές, η έξοδος είναι κοντά στο μηδέν (όχι μεταβολή).
- Σε περιοχές με έντονες μεταβολές (π.χ. ακμές), τονίζει τις διαφορές.
- Αν έχει κατευθυντικότητα, ενισχύει ιδιαίτερα τις ακμές στη συγκεκριμένη διεύθυνση.