

# Εκπαίδευση Μοντέλου PIDNet για Ανίχνευση Πυρκαγιάς

## 1. Εισαγωγή και Στόχος

Σκοπός του παρόντος έργου είναι η χρήση και εκπαίδευση του μοντέλου **PIDNet** για την επίλυση του προβλήματος της Σημασιολογικής Κατάτμησης (Semantic Segmentation) σε περιβάλλοντα πυρκαγιάς.

## 2. Προετοιμασία Δεδομένων (Data Engineering)

Η διαδικασία προετοιμασίας των δεδομένων αποτέλεσε την βάση του project, καθώς έπρεπε να μετατρέψουμε ακατέργαστα δεδομένα από εξωτερικό εργαλείο σε μορφή συμβατή με το PIDNet (CamVid format).

### 2.1. Καθορισμός Κλάσεων

Οι κλάσεις που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου προέκυψαν απευθείας από τα labels και τις κλάσεις που δημιουργήθηκαν κατά τη διαδικασία του annotation (επισημείωση).

Συγκεκριμένα, κατά τη φάση της επισημείωσης των εικόνων, χρησιμοποιήθηκαν μάσκες (masks) για να οριοθετηθούν οι περιοχές ενδιαφέροντος. Από αυτή τη διαδικασία προέκυψαν οι εξής **4 κλάσεις**:

- Fire (Φωτιά)**
- Smoke (Καπνός)**
- Burnt (Καμμένο)**
- Background (Υπόβαθρο)** - (Η κλάση που περιλαμβάνει οτιδήποτε δεν ανήκει στα παραπάνω).

### 2.2. Ροή Εργασίας: Από το CVAT στο Project Structure

Η επισημείωση των εικόνων πραγματοποιήθηκε στο εργαλείο **CVAT**. Η διαδικασία μεταφοράς των δεδομένων στο περιβάλλον του PIDNet έγινε ως εξής:

- Εξαγωγή (Export):** Τα δεδομένα εξήχθησαν από το CVAT επιλέγοντας το format "CamVid 1.1".
- Δομική Ασυμβατότητα:** Το zip αρχείο που προέκυψε από το CVAT είχε μια δομή φακέλων (default, defaultannot) η οποία δεν ήταν συμβατή με τον κώδικα του PIDNet.
- Αναδιάρθρωση (Restructuring):** Έγινε αναδιάρθρωση των αρχείων για να δημιουργηθεί μια "επίπεδη" (flat) δομή που απαιτεί το μοντέλο. Μεταφέρθηκαν τα αρχεία στους εξής φακέλους εντός του project:

- data/camvid/images/: Όλες οι εικόνες εισόδου.
- data/camvid/labels/: Όλες οι μάσκες.

### 2.3. Χαρτογράφηση Δεδομένων (List Files)

Το μοντέλο βασίζεται σε αρχεία κειμένου (.lst) που περιέχουν τα ονόματα εικόνων και масκών. Έγινε επεξεργασία στα αρχεία train.lst, val.lst και trainval.lst ώστε να περιέχουν τα ονόματα και τους συνδυασμούς εικόνων-labels του καινούριου dataset.

## 3. Παραμετροποίηση Μοντέλου (Configuration)

Για να λειτουργήσει το μοντέλο με τα δικά μας δεδομένα και το συγκεκριμένο hardware (Kaggle T4 GPUs), έγιναν τροποποιήσεις στο αρχείο ρυθμίσεων **configs/camvid/pidnet\_small\_camvid.yaml**.

Όλες οι αλλαγές στο αρχείο έχουν σημειωθεί με το σχόλιο **# --- ΑΛΛΑΓΗ ΕΔΩ**, ώστε να είναι εύκολα εντοπίσιμες.

Ένα χαρακτηριστικό παράδειγμα αλλαγής αφορά τον ορισμό των κλάσεων και του Dataset:

```
YAML
DATASET:
  # --- ΑΛΛΑΓΗ ΕΔΩ
  DATASET: camvid
  # --- ΑΛΛΑΓΗ ΕΔΩ
  NUM_CLASSES: 4
  ROOT: data/
```

Εκτός από το παραπάνω, πραγματοποιήθηκαν οι εξής βασικές ρυθμίσεις (τις οποίες μπορείτε να δείτε αναλυτικά αναζητώντας το σχετικό σχόλιο στον κώδικα):

- Ενεργοποίηση παράλληλης εκπαίδευσης και στις **2 κάρτες γραφικών (GPUs)**.
- Απενεργοποίηση των **προ-εκπαιδευμένων βαρών (Pretrained weights)**, ώστε το μοντέλο να μάθει αποκλειστικά από τα δικά μας δεδομένα.
- Απλοποίηση της συνάρτησης κόστους (Loss Function) με απενεργοποίηση του OHEM και του Boundary Loss για μεγαλύτερη σταθερότητα στην αρχή.
- Απενεργοποίηση του CUDNN Benchmark για αποφυγή σφαλμάτων μνήμης.

## 4. Τεχνικές Βελτιώσεις και Διορθώσεις (Stability Fixes)

Κατά τη διάρκεια της υλοποίησης, υπήρξαν τεχνικά εμπόδια που αφορούσαν τη συμβατότητα των δεδομένων και τη σταθερότητα της εκπαίδευσης. Εφαρμόστηκαν οι εξής λύσεις:

**1. Ομοιομορφία Εικόνων (Resize)** Το σύστημα αδυνατούσε να επεξεργαστεί ταυτόχρονα (batch processing) εικόνες που είχαν διαφορετικές διαστάσεις. Χρησιμοποιήθηκε μηχανισμός

που προσαρμόζει αυτόματα όλες τις εικόνες εισόδου σε συγκεκριμένη ανάλυση (**960x720**) πριν αυτές εισέλθουν στο δίκτυο, εξασφαλίζοντας την ομαλή ροή των δεδομένων.

**2. Διαχείριση Μνήμης (Memory Safety)** Παρατηρήθηκε αστάθεια κατά τη μεταφορά δεδομένων από τον επεξεργαστή στην κάρτα γραφικών, η οποία οδηγούσε σε κατάρρευση της μνήμης (crashes). Βελτιστοποιήθηκε ο τρόπος φόρτωσης (Data Loading), μετατρέποντας τα δεδομένα στην κατάλληλη μορφή (Tensors) νωρίτερα στη διαδικασία, κάτι που έλυσε οριστικά τα θέματα μνήμης στο περιβάλλον του Kaggle.

**3. Σφάλμα Learning Rate** Διορθώθηκε ένα υπολογιστικό ζήτημα όπου το Learning Rate μηδενιζόταν τελείως στο τέλος της εκπαίδευσης, προκαλώντας διακοπή του αλγορίθμου. Προσθέσαμε μια ελάχιστη τιμή ασφαλείας ώστε η εκπαίδευση να ολοκληρώνεται ομαλά μέχρι την τελευταία εποχή.

**4. Ασφάλεια Ετικετών (Label Safety)** Για να προστατευθεί η εκπαίδευση από τυχόν "θόρυβο" στα δεδομένα (π.χ. pixels με λανθασμένες τιμές κλάσεων), προστέθηκε ένας μηχανισμός ασφαλείας. Ο μηχανισμός αυτός εντοπίζει αυτόματα τα προβληματικά pixels και τα αγνοεί κατά τον υπολογισμό του σφάλματος, επιτρέποντας στο μοντέλο να συνεχίσει να μαθαίνει χωρίς διακοπές.

## 5. Διαδικασία Εκπαίδευσης και Επαλήθευσης (Training & Validation)

Η ανάπτυξη του μοντέλου υλοποιήθηκε μέσω μιας επαναληπτικής διαδικασίας εντοπισμού σφαλμάτων και βελτιστοποίησης. Η εκπαίδευση πραγματοποιήθηκε στο Kaggle γιατί προσφέρει χρήση 2 NVIDIA Tesla T4 GPUs για αρκετές ώρες.

### 5.1. Παράμετροι Εκπαίδευσης

Για την εκπαίδευση του δικτύου ορίστηκαν οι ακόλουθες παράμετροι από configurations:

- **Διάρκεια (Epochs):** Η διαδικασία ολοκληρώθηκε στις **120/200 εποχές**. Επιλέχθηκε αυτό το σημείο διακοπής καθώς το μοντέλο είχε επιτύχει σταθεροποίηση αποτελεσμάτων και δεν βελτιωνόταν άλλο η πρόβλεψη.
- **Μέγεθος Δέσμης (Batch Size):** 8 εικόνες ανά βήμα (4 ανά GPU), για βέλτιστη διαχείριση της μνήμης VRAM.
- **Βελτιστοποιητής (Optimizer):** Χρησιμοποιήθηκε ο αλγόριθμος **SGD** (Stochastic Gradient Descent) με Momentum, ξεκινώντας με ρυθμό μάθησης (Learning Rate) **0.01** και εφαρμόζοντας πολυωνυμική μείωση (polynomial decay) κατά τη διάρκεια της εκπαίδευσης.
- **Συναρτήσεις Κόστους (Loss Functions):** Χρησιμοποιήθηκε συνδυασμός Cross Entropy Loss (για την ταξινόμηση των pixels) και Boundary Loss (για την έμφαση στα όρια των αντικειμένων).

### 5.2. Αντιμετώπιση Προβλημάτων Δεδομένων

Κατά τα αρχικά στάδια της εκπαίδευσης, παρατηρήθηκε το φαινόμενο της "**Κατάρρευσης Μοντέλου**" (**Model Collapse**), όπου το δίκτυο προέβλεπε μία ενιαία κλάση για ολόκληρη την εικόνα, παρόλο που το σφάλμα (Loss) φαινόταν να μειώνεται.

Το σφάλμα ήταν στον τρόπο που είχαν εξαχθεί οι μάσκες (Labels) από το εργαλείο επισημείωσης (CVAT). Οι μάσκες δεν περιείχαν τα αναμενόμενα Class IDs (0, 1, 2, 3), αλλά τιμές χρωματισμού RGB που μεταφράστηκαν στην συνέχεια σε τιμές φωτεινότητας (grayscale values) των χρωμάτων (π.χ. το κόκκινο της φωτιάς είχε τιμή 76 και το μπλε του καπνού τιμή 24). Αυτό είχε ως αποτέλεσμα το δίκτυο να αγνοεί αυτές τις περιοχές ως "θόρυβο".

Γι' αυτό αναπτύχθηκε ένας αλγόριθμος που σαρώνει κάθε μάσκα πριν την εισαγωγή της στο δίκτυο και μετατρέπει τις τιμές φωτεινότητας στα σωστά IDs:

- 76 -> Class 0 (Φωτιά)
- 24 -> Class 1 (Καπνός)
- 147 -> Class 2 (Καμμένα)

Μετά την εφαρμογή αυτής της διόρθωσης, η εκπαίδευση προχώρησε κανονικά, με την ακρίβεια να αυξάνεται σταθερά.

## 6. Αποτελέσματα

Η αξιολόγηση του μοντέλου έγινε τόσο σε ποσοτικό επίπεδο (μετρικές) όσο και σε ποιοτικό (οπτικός έλεγχος σε άγνωστες εικόνες).

### 6.1. Ποσοτικά Αποτελέσματα

Στην εποχή 120, το μοντέλο σημείωσε τις εξής επιδόσεις στο σύνολο επαλήθευσης (Validation Set):

- **Συνολική Ακρίβεια Pixels (Pixel Accuracy): 80.6%**
- **IoU (Intersection over Union) Καπνού: 77.8%**
  - Η τιμή αυτή θεωρείται εξαιρετικά υψηλή για το συγκεκριμένο πρόβλημα, δεδομένης της ημιδιαφανούς φύσης του καπνού και της δυσκολίας καθορισμού των ορίων του. Αποδεικνύει ότι το μοντέλο έμαθε επιτυχώς τα μορφολογικά χαρακτηριστικά του καπνού.
- **Φωτιά, Καμμένα**
  - Η ακρίβεια είναι πολύ χαμηλή και στην ουσία το μοντέλο αποτυγχάνει. Για παράδειγμα παρατηρείται ότι κάποια edges το μπερδεύουν και θεωρεί φωτιά όταν υπάρχει μεγάλη διαφορά χρώματος/φωτεινότητας από καπνό σε μπλε ουρανό. Χρειάζονται περισσότερες εικόνες με φωτιά και καμμένα για training.

## 7. Συμπεράσματα

Η πειραματική διαδικασία κατέδειξε ότι το μοντέλο είναι **πλήρως λειτουργικό για την ανίχνευση καπνού**, πετυχαίνοντας υψηλή ακρίβεια που το καθιστά κατάλληλο ως σύστημα έγκαιρης προειδοποίησης (καθώς ο καπνός είναι συνήθως το πρώτο ορατό ίχνος μιας πυρκαγιάς). Τα σφάλματα που αφορούν τη σύγχυση φωτιάς/έντονου φωτός είναι αναμενόμενα σε αυτό το στάδιο εκπαίδευσης και μπορούν να εξαλειφθούν με την επέκταση του συνόλου δεδομένων και τη συνέχιση της εκπαίδευσης για περισσότερες εποχές.

