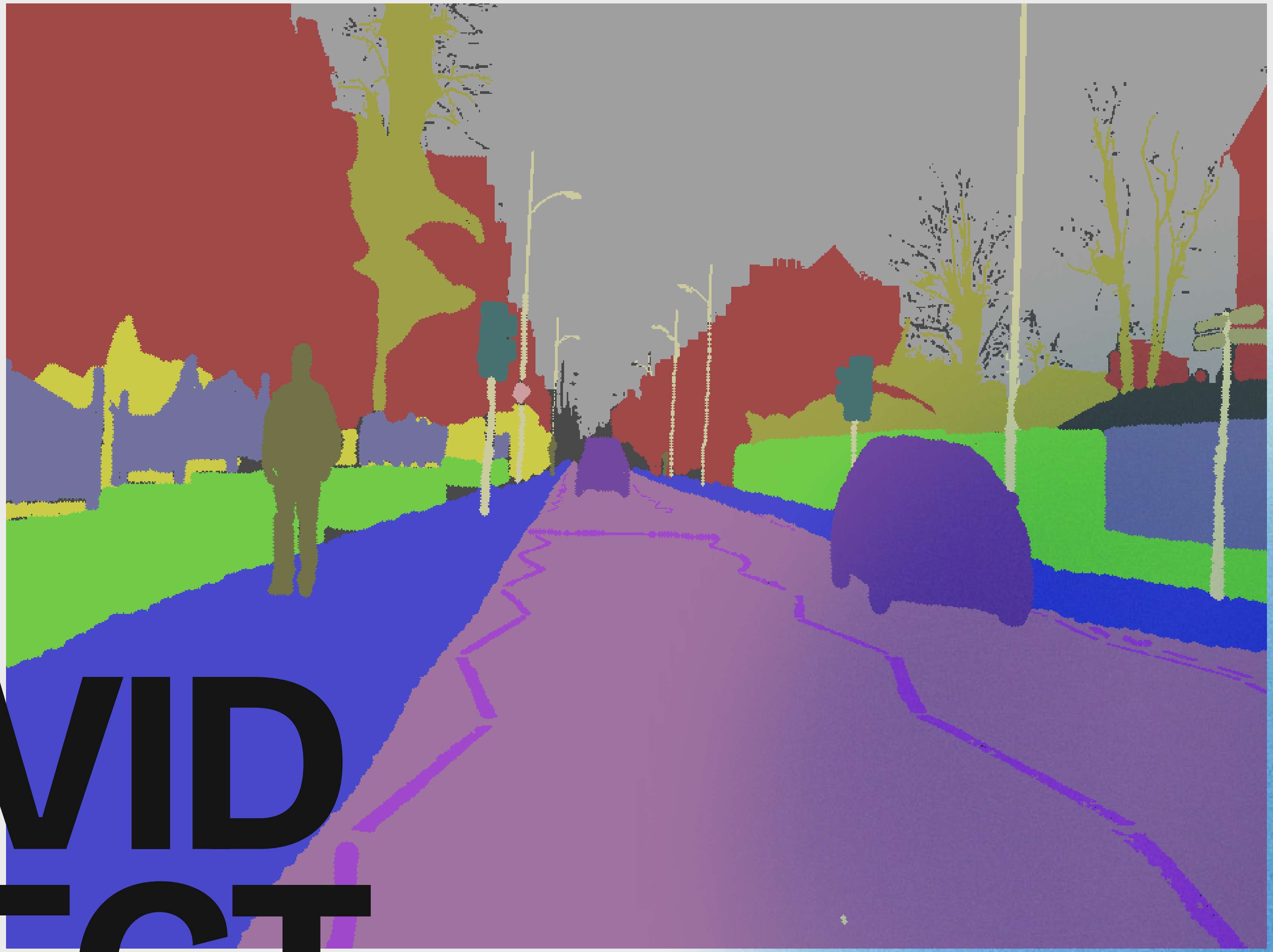
BORGIA MATTIA
PALESTRA LEONARDO



CAM VID PROJECT

SEMANTIC SEGMENTATION

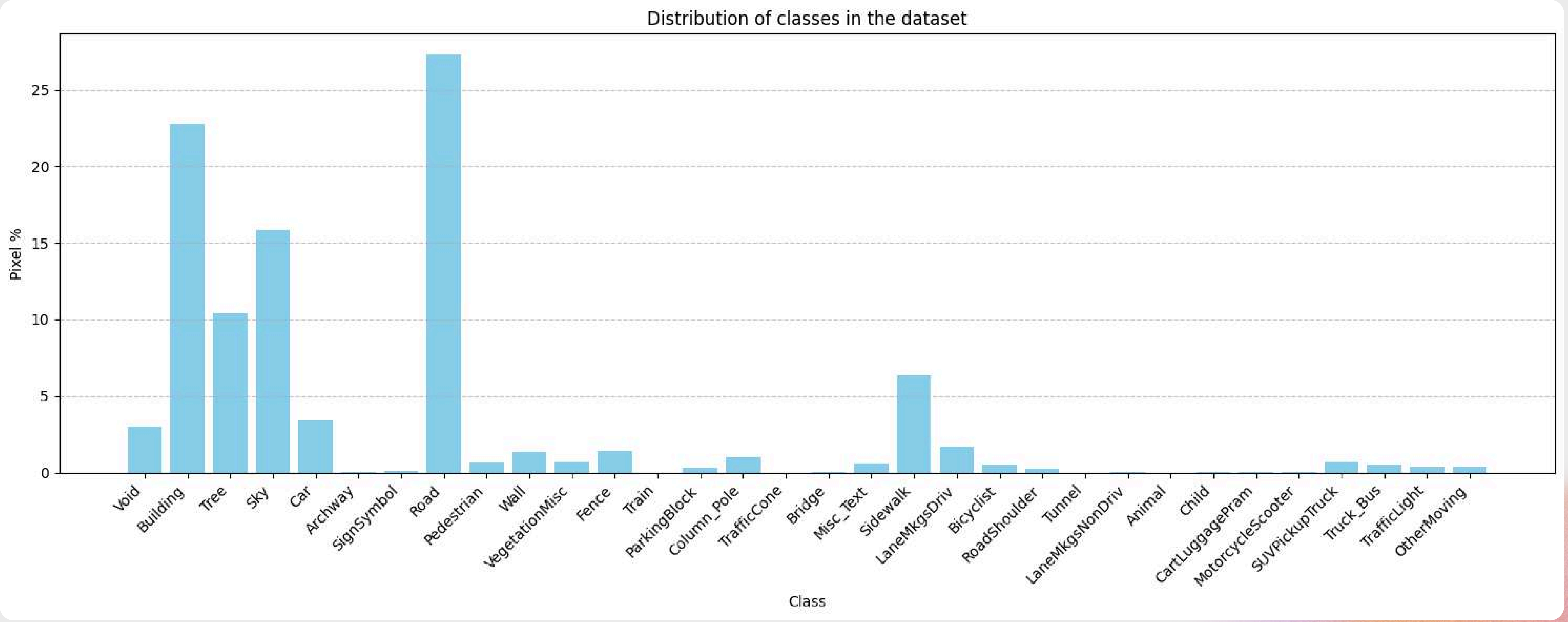
COMPUTER VISION TECHNIQUE THAT CLASSIFIES EACH PIXEL IN AN IMAGE INTO A PREDEFINED CATEGORY, ALLOWING FOR PRECISE IDENTIFICATION AND LOCALIZATION OF OBJECTS OR REGIONS BASED ON THEIR CLASS.

DATASET - CAMVID

701 IMAGES
701 MASKS

720X960

32 CLASSES



DATASET - CAMVID - PREPROCESSING

11 CLASSES

SKY	SKY
BUILDING	BUILDING, BRIDGE, TUNNEL, GARAGE
POLE	POLE, TRAFFIC LIGHT, TRAFFIC SIGN
ROAD	ROAD, LANE MARKINGS
PAVEMENT	SIDEWALK, GROUND, PARKING
TREE	TREE, VEGETATION, BUSH, PLANT
SIGN/SYMBOL	BILLBOARD, SIGN SYMBOL, TRAFFIC SIGN, TRAFFIC LIGHT
FENCE	FENCE, WALL, RAILING, BARRIER
CAR	CAR, TRUCK, BUS, TRAIN, MOTORCYCLE, BICYCLE
PEDESTRIAN	PEDESTRIAN, RIDER
BICYCLIST	CYCLIST, BICYCLE

PRO

- **BETTER CLASS BALANCE**
- **BETTER GENERALIZATION**
- **EASIER TO COMPARE**
- **FOCUS ON MEANINGFUL CATEGORIES**

DATASET - CAMVID - PREPROCESSING

CONVERTED THE MASKS FROM PNG TO NUMPY

RESIZED FROM 720X960 TO 512X512

NORMALIZE PIXEL IN A RANGE OF [0,1] BY DIVIDING EACH PIXEL VALUE BY 255

SPLIT THE DATASET IN
- TRAINING **367** OBSERVATIONS
- VALIDATION **101** OBSERVATIONS
- TEST **233** OBSERVATIONS

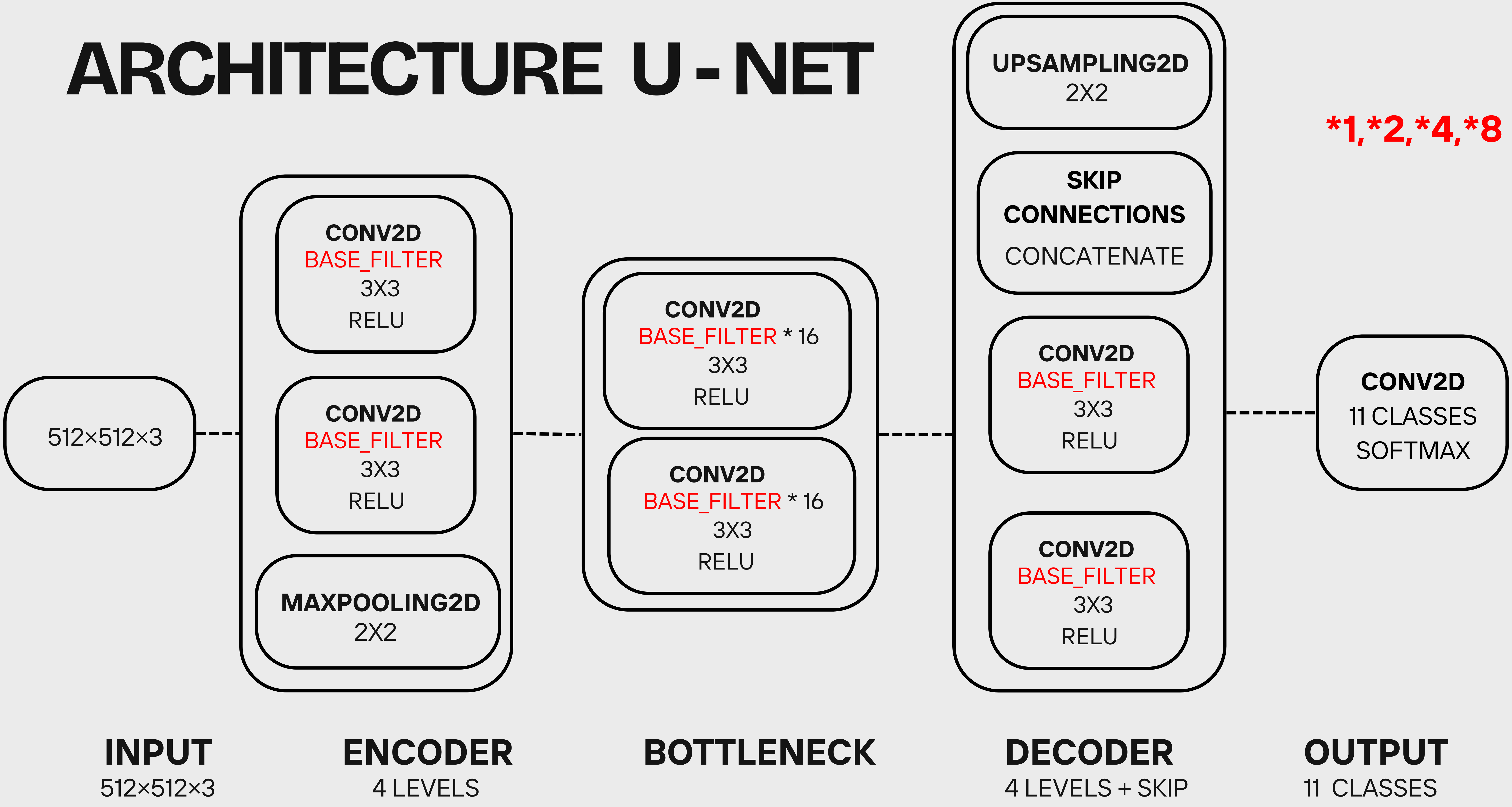
DATASET - CAMVID - PREPROCESSING

ONLINE DATA AUGMENTATION

PROCESS OF APPLYING DATA AUGMENTATION TRANSFORMATIONS IN REAL-TIME DURING MODEL TRAINING, RATHER THAN GENERATING AND STORING THE AUGMENTED DATA BEFOREHAND.
EACH TRAINING BATCH IS AUGMENTED ON THE FLY, WHICH SAVES STORAGE SPACE AND PROVIDES A VIRTUALLY INFINITE VARIETY OF TRAINING EXAMPLES.

- ZOOM-IN USING RANDOM CROP
- BRIGHTNESS ADJUSTMENT
- CONTRAST ADJUSTMENT
- SATURATION ADJUSTMENT
- HUE ADJUSTMENT
- GAUSSIAN NOISE

ARCHITECTURE U-NET



SELECTION BEST LOSS AND OPTIMIZER

LOSS
- SPARSE CATEGORICAL CROSSENTROPY
- DICE LOSS
- DICE CE LOSS

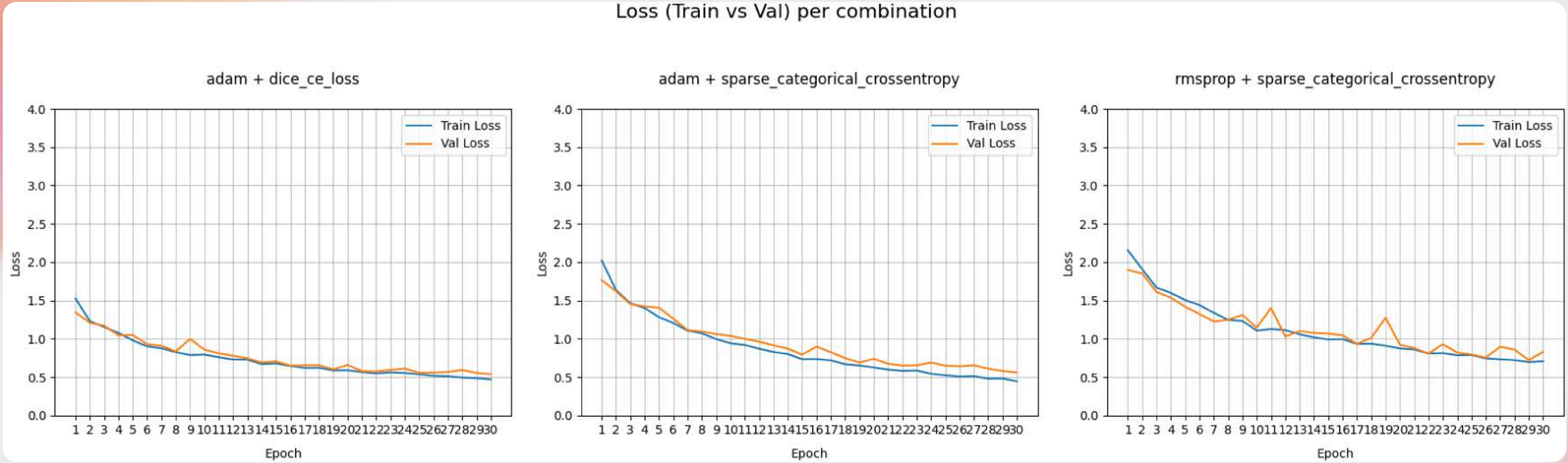
OPTIMIZER
- ADAM
- RMSPROP

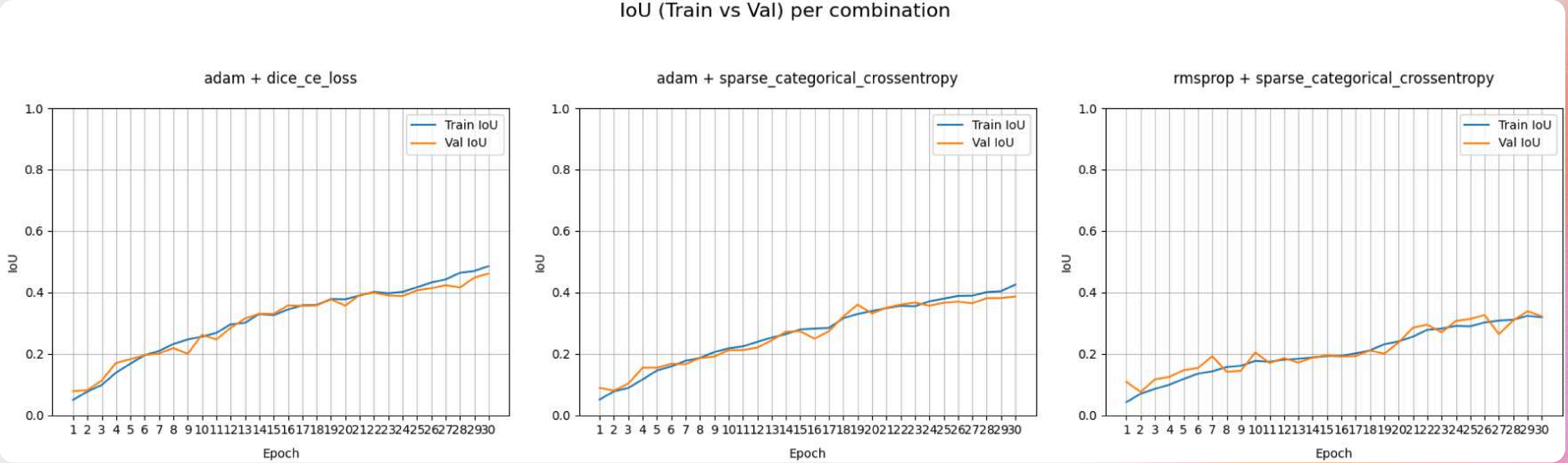
6 COMBINATIONS OF LOSS AND OPTIMIZER
SORTED BY IOU (INTERSECTION OVER UNION)

#1
OPTIMIZER: **ADAM**
LOSS FUNCTION:
DICE_CE_LOSS
VALIDATION IOU: **0.4616**
FINAL VAL ACCURACY: **0.8523**
FINAL VAL LOSS: **0.5379**
FINAL TRAIN IOU: **0.4857**

#2
OPTIMIZER: **ADAM**
LOSS FUNCTION:
SPARSE_CATEGORICAL_CROSSENTROPY
VALIDATION IOU: **0.3865**
FINAL VAL ACCURACY: **0.8424**
FINAL VAL LOSS: **0.5617**
FINAL TRAIN IOU: **0.4254**

#3
OPTIMIZER: **RMSPROP**
LOSS FUNCTION:
SPARSE_CATEGORICAL_CROSSENTROPY
VALIDATION IOU: **0.3224**
FINAL VAL ACCURACY: **0.7478**
FINAL VAL LOSS: **0.8293**
FINAL TRAIN IOU: **0.3192**

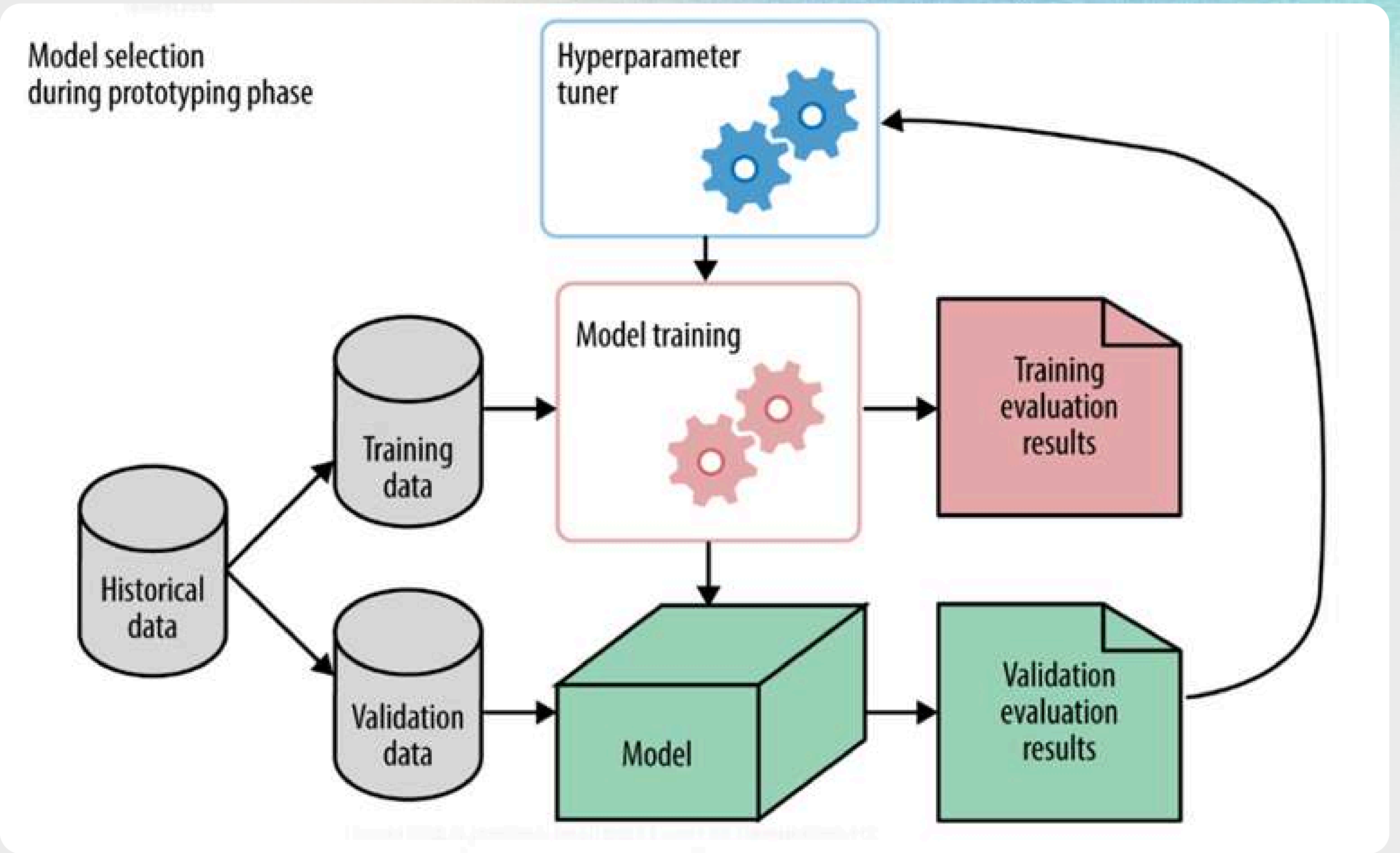

Loss (Train vs Val) per combination
adam + dice_ce_loss
adam + sparse_categorical_crossentropy
rmsprop + sparse_categorical_crossentropy
Train Loss
Val Loss
Loss
Epoch


IoU (Train vs Val) per combination
adam + dice_ce_loss
adam + sparse_categorical_crossentropy
rmsprop + sparse_categorical_crossentropy
Train IoU
Val IoU
IoU
Epoch

TUNE HYPERPARAMETERS

HYPERPARAMETER TUNING
BASE FILTERS: [8, 16, 32]
LEARNING RATE: [0.01, 0.001, 0.0001]
ALPHA = [0.2 - 0.8, STEP = 0.1] *
SMOOTH = [0.2 - 0.8, STEP = 0.1] **

IF **DICE_CE_LOSS** SELECTED*
IF **DICE_LOSS** SELECTED**



MORE INFO:

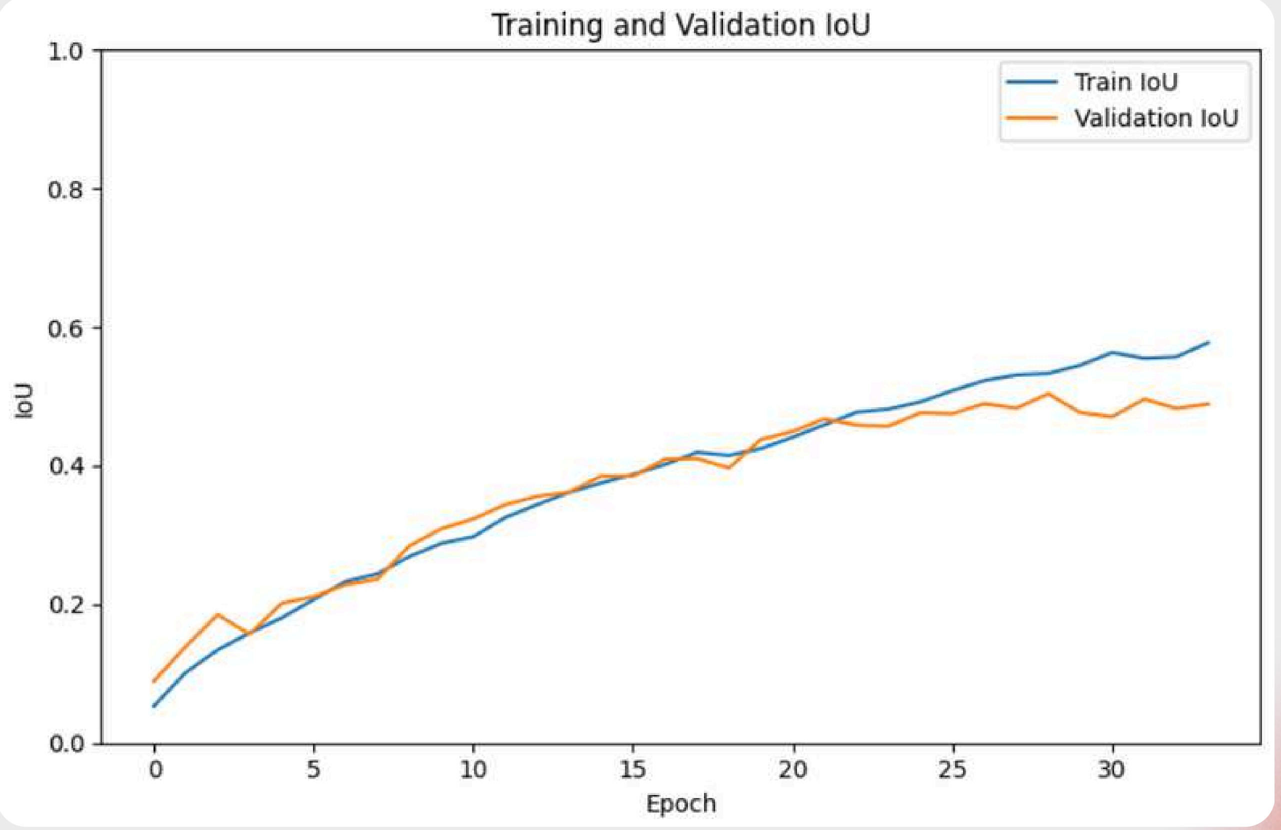
TUNING STRATEGY: KERAS TUNER — HYPERBAND
OBJECTIVE: MAXIMIZE VALIDATION IOU (VAL_IOU)
MAX EPOCHS PER TRIAL: 20
FACTOR: 3
PATIENCE: 5 EPOCHS

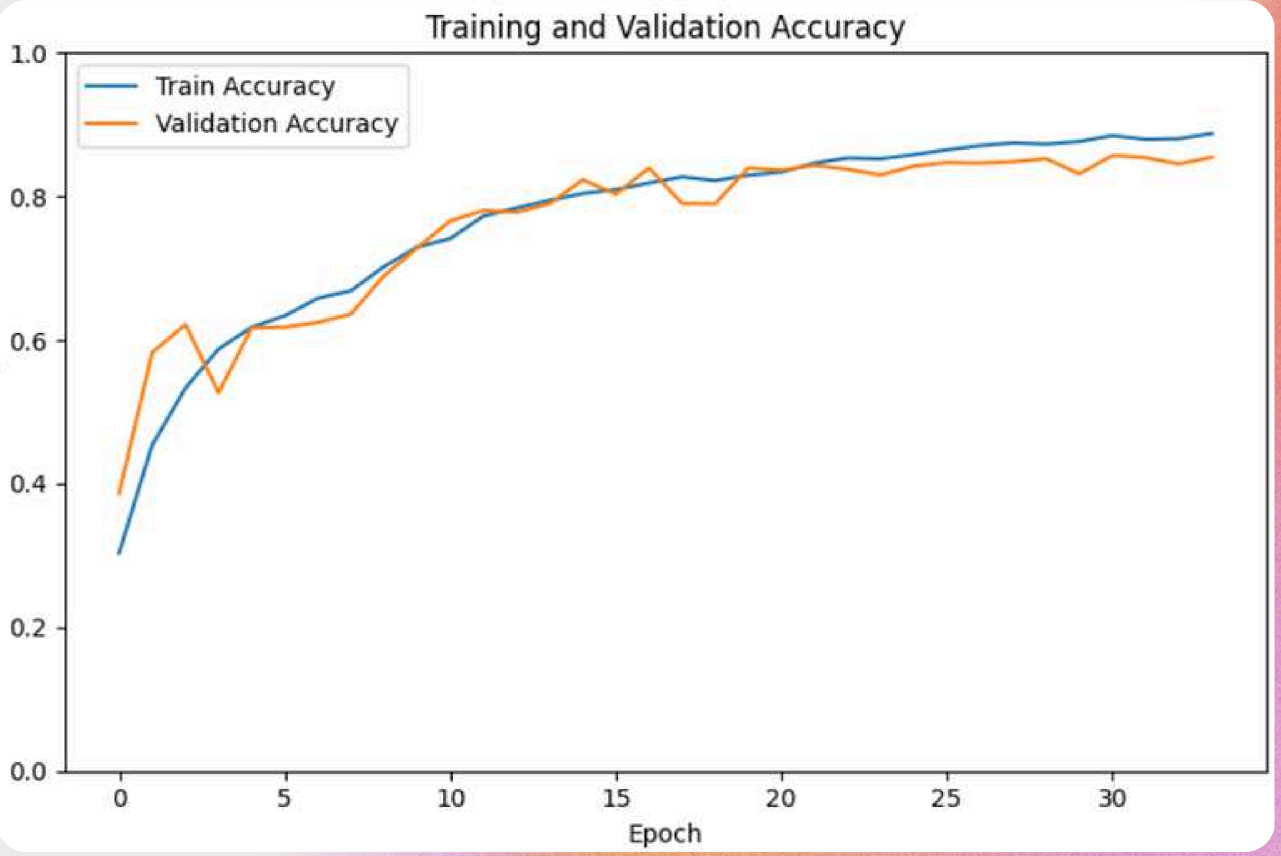
BEST HYPERPARAMETERS FOUND:
LOSS: DICE_CE_LOSS, **OPTIMIZER**: ADAM, **BASE FILTERS**: 32, **LEARNING RATE**: 0.001, **ALPHA**: 0.2

RESULTS

THESE ARE THE FINAL RESULTS OF THE MODEL ON THE TEST SET:
IOU: 0.4447**—** **ACCURACY:** 0.8062**—** **LOSS:** 0.5849.

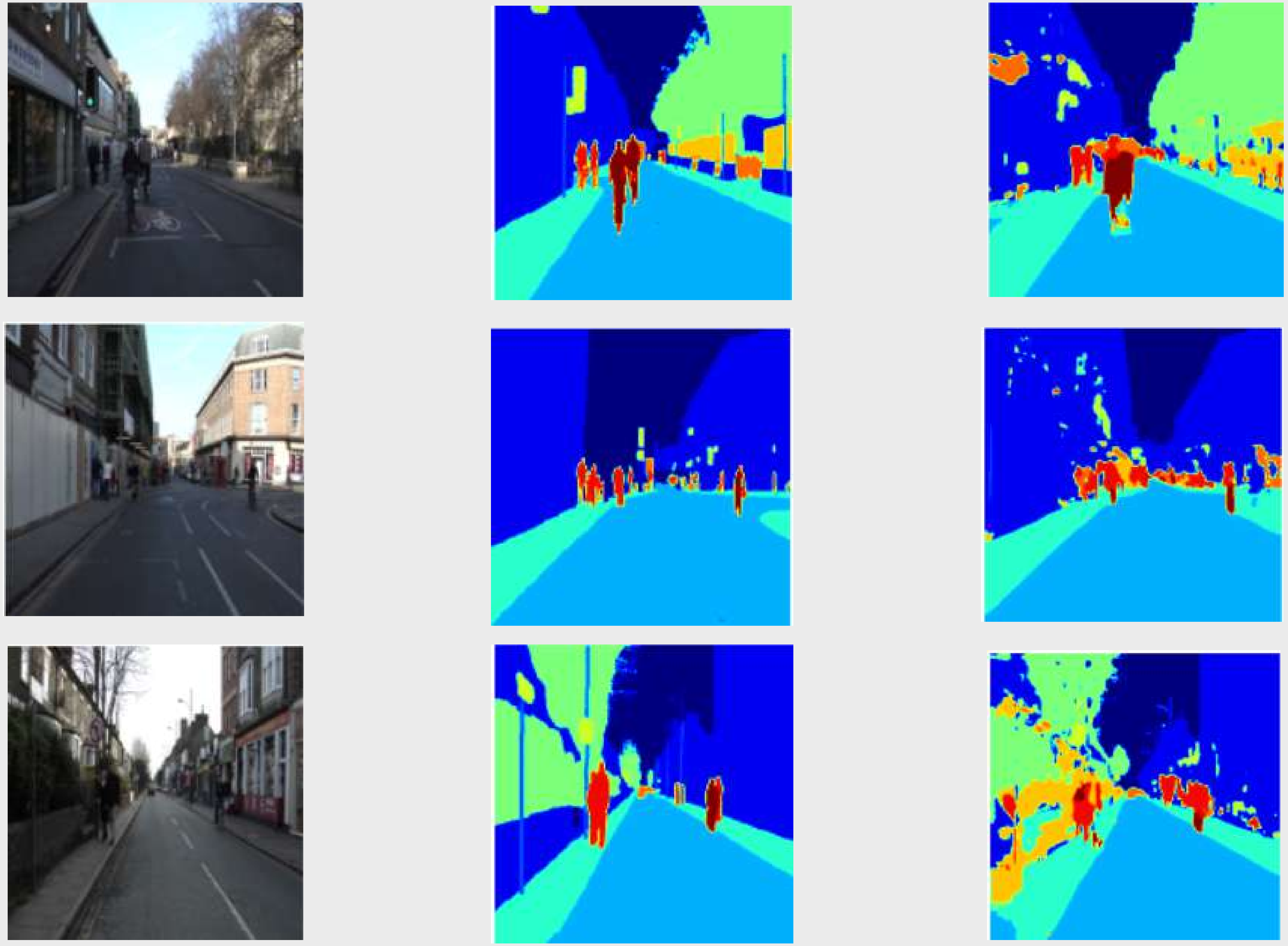
IOU PER CLASS (SORTED):
CLASS 3 (ROAD): IOU = 0.9163
CLASS 0 (SKY): IOU = 0.8722
CLASS 4 (PAVEMENT): IOU = 0.7824
CLASS 5 (TREE): IOU = 0.7789
CLASS 1 (BUILDING): IOU = 0.7551
CLASS 8 (CAR): IOU = 0.7056
CLASS 10 (BICYCLIST): IOU = 0.6649
CLASS 9 (PEDESTRIAN): IOU = 0.5794
CLASS 7 (FENCE): IOU = 0.5706
CLASS 2 (POLE): IOU = 0.5594
CLASS 6 (SIGN/SYMBOL): IOU = 0.5392





RESULTS

REAL | GROUND TRUTH | PREDICTED



ADDITIONAL TRIALS

OFFLINE DATA AUGMENTATION
IN ADDITION TO ONLINE DATA AUGMENTATION

CONV2D TRANSPOSE
INSTEAD OF UPSAMPLING

INSTANCE NORMALIZATION
AFTER EACH CONVOLUTIONAL LAYER

SPATIAL DROPOUT
IN EACH BLOCK

ATTENTION GATES
BEFORE CONCATENATION

WHY IT DIDN’T PERFORM BETTER?

- EXCESSIVE COMPLEXITY WITH FEW DATA
- NEED MORE EPOCHS TO FULLY CONVERGE
- OPTIMIZATION PROCESS MORE CHALLENGING AND SENSITIVE TO HYPERPARAMETERS.
- EXCESSIVE DROPOUT OR UNREALISTIC OFFLINE AUGMENTATION MADE LOSE THE POWER OF GENERALIZATION

NEED FURTHER ANALYSIS

REFERENCES

LONG, J., SHELHAMER, E., & DARRELL, T. (2015). **FULLY CONVOLUTIONAL NETWORKS FOR SEMANTIC SEGMENTATION.**

RONNEBERGER, O., FISCHER, P., & BROX, T. (2015). U-NET: **CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE**

ZHANG, Y., YANG, R., WANG, J., CHEN, N., & DAI, Q. (2021). **THE IMPACT OF PARAMETERS ON SEMANTIC SEGMENTATION**

ERIŞEN, S. (2024). SERNET-FORMER: **SEMANTIC SEGMENTATION BY EFFICIENT RESIDUAL NETWORK WITH ATTENTION-BOOSTING GATES AND ATTENTION-FUSION NETWORKS**.

THANK YOU
QUESTIONS?