ROAD SEGMENTATION WITH NEURAL NETWORKS

REVOLUTIONIZING TRANSPORTATION AND URBAN PLANNING

TODAY'S

INTRODUCTION
DEEP LEARNING MODELS
EXPERIMENTAL SETUP & RESULTS
ANALYSIS OF MODELS
CONCLUSION

AGENDA

What is image Segmentation ?

Image segmentation in neural networks is a computer vision technique that uses complex neural networks to divide an image into multiple regions or objects based on their characteristics. The goal is to simplify and make the image easier to analyze. In image segmentation, pixels are grouped based on their similarity in color, intensity, texture, or other characteristics. The output of this annotation is a segmentation mask, which shows the boundary and shape of each class in the image.

What is Road Segmentation ?

A crucial technology that enables vehicles to perceive and understand their surroundings.

It acts as a visual perception system, allowing vehicles to identify and classify road elements, such as lanes, crosswalks, and obstacles.

WHY SEGMENTATION
FOR AUTNOMOUS VEHICLES

Perception and Understanding :

Road segmentation helps autonomous vehicles understand the driving environment, make informed decisions, plan trajectories, and ensure safe navigation.

Lane Keeping and Trajectory Planning :

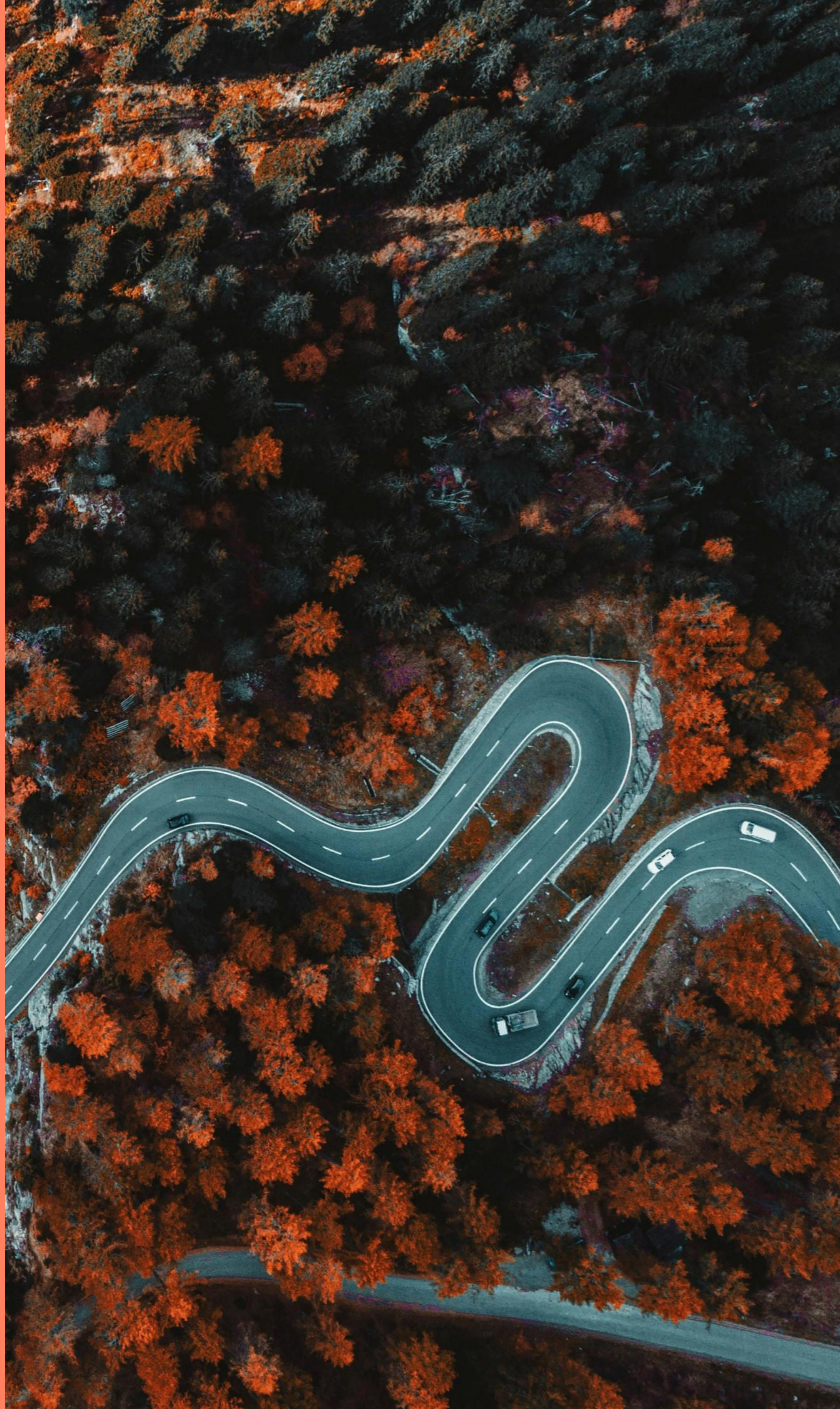
Lane detection enables proper positioning, smooth maneuvers, trajectory planning, and anticipation of upcoming road features like curves or intersections.

Obstacle Detection and Avoidance :

Distinguishing road surface from hazards like vehicles, pedestrians, and debris helps maintain safe distance and navigate complex environments.

Compliance with Traffic Rules and Regulations :

Identifying traffic signs, signals, and markings allows vehicles to adjust behavior, follow rules, and integrate into transportation systems.



Segmentation and Urban Planning

Road segmentation revolutionizes urban planning for smarter, safer, sustainable transportation by leveraging data to assess infrastructure, optimize traffic, and enhance livability.

Infrastructure Assessment & Maintenance

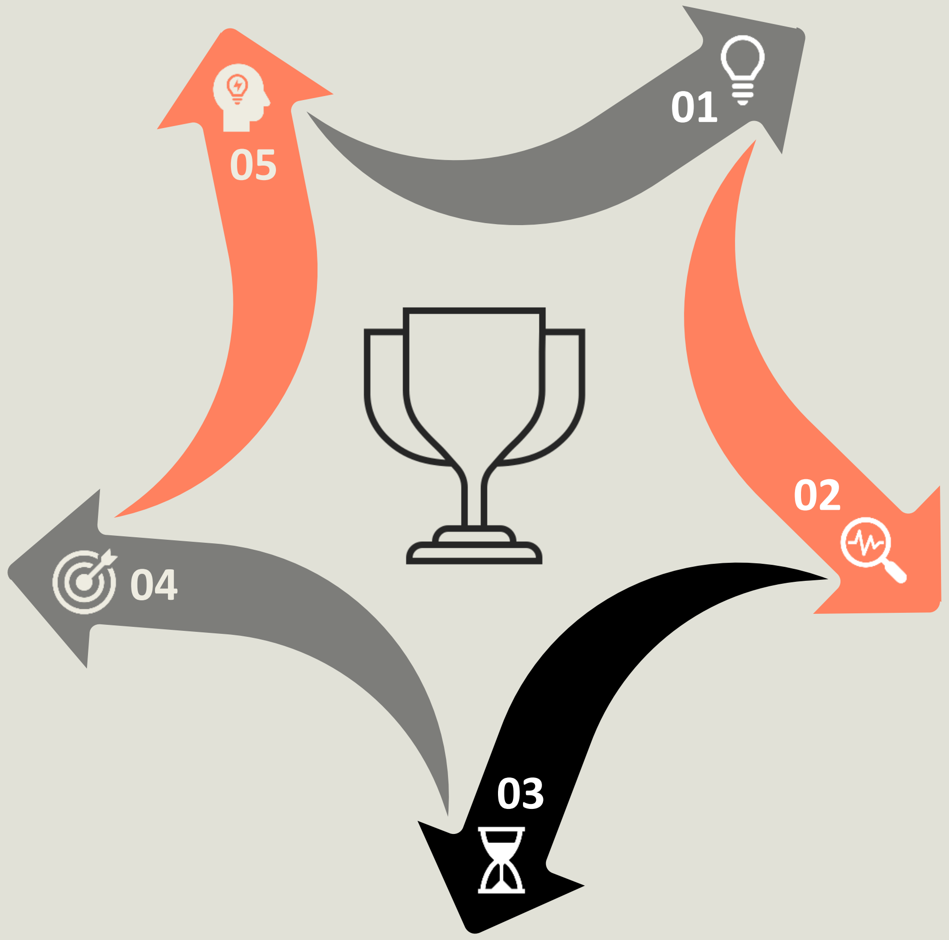
Segmented road data helps identify areas needing maintenance, repair, or upgrades, prioritizing projects and allocating resources efficiently

Traffic Flow Analysis & Optimization

Understanding road segment utilization optimizes traffic management, signal timings, calming measures, and efficient networks to reduce congestion.

Traffic Flow Analysis & Optimization

Analyzing segmented data locates high-risk zones for pedestrians and cyclists, guiding safety measures like crossings and bike lanes.



Sustainable Transportation Planning

Identifying suitable road segments for bus lanes, bike paths, or pedestrian-friendly streets promotes sustainable transportation alternatives.

Land Use & Zoning Decisions

Understanding road segment characteristics and usage informs compatible land use and zoning decisions for residential, commercial, or industrial areas.

Deep Learning for Semantic Segmentation

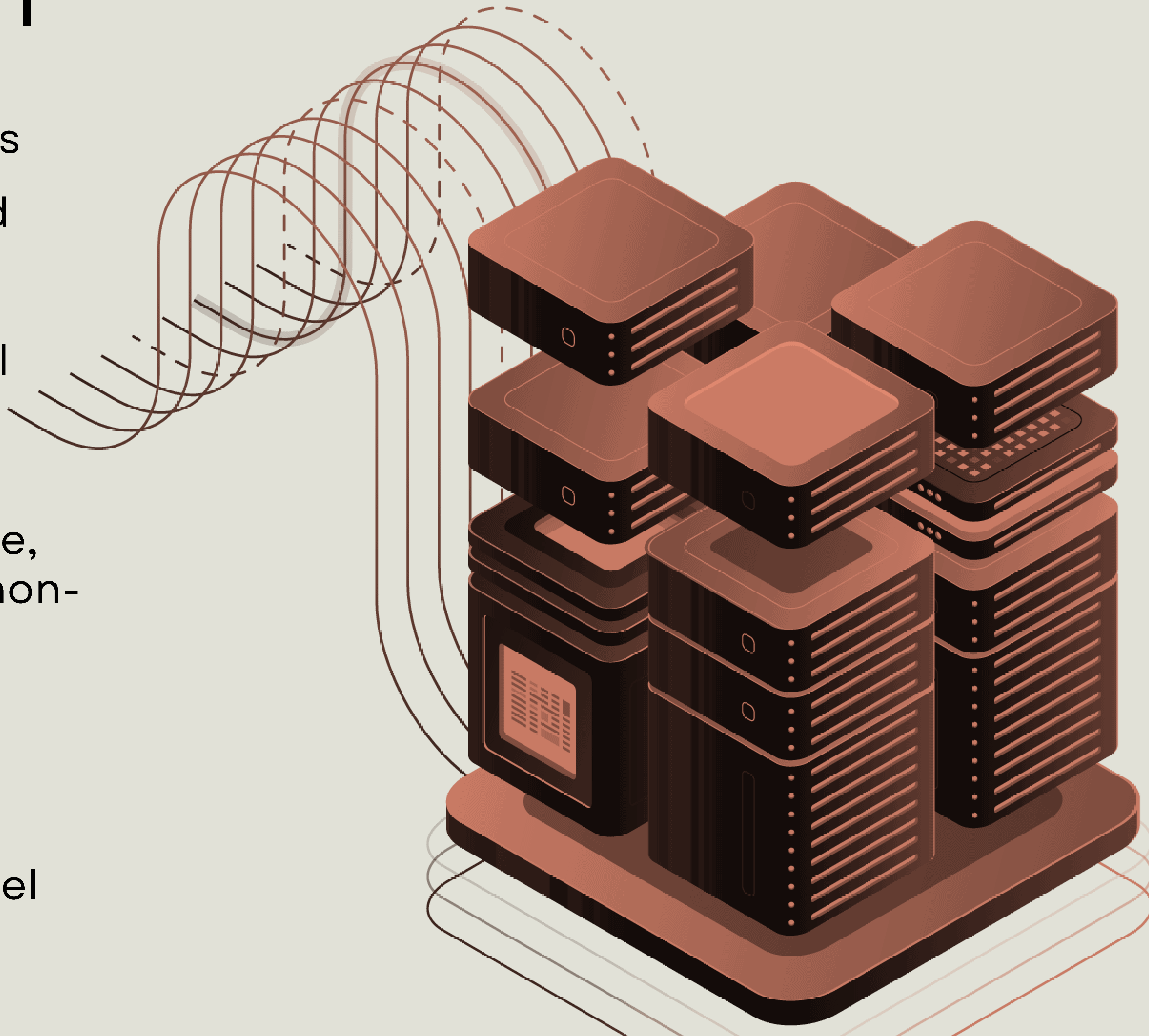
And popular architectures

Dataset Description

- A collection of high-resolution images and videos depicting roads, captured from various viewpoints, angles, and lighting conditions, in urban and rural environments.

- Pixel-level annotations for each image, labeling each pixel as either road or non-road, provided as masks or overlays.

- Images covering a wide range of environmental conditions, such as different times of day, weather conditions, and seasons, ensure model robustness.



Processing Steps

Datasets

Prepare Datasets

Training Data

Validation Data

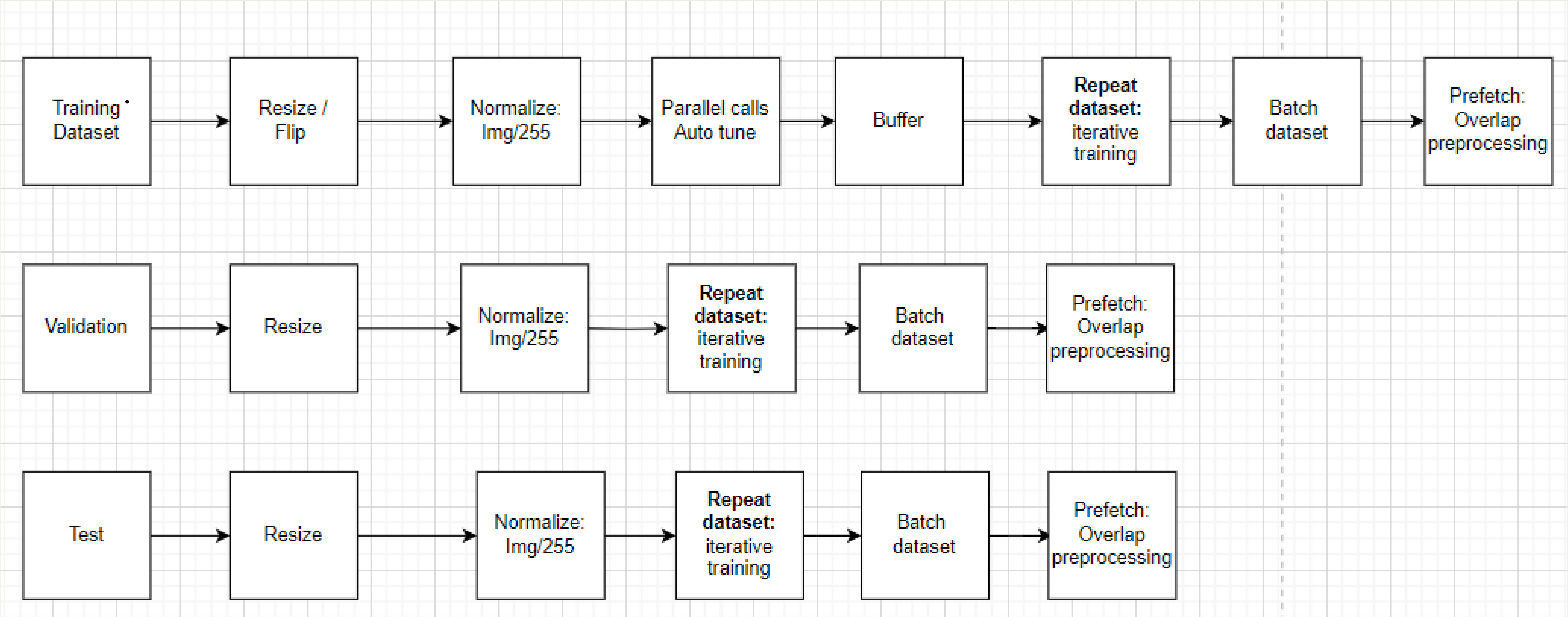
Testing Data

1000 Buffer Size

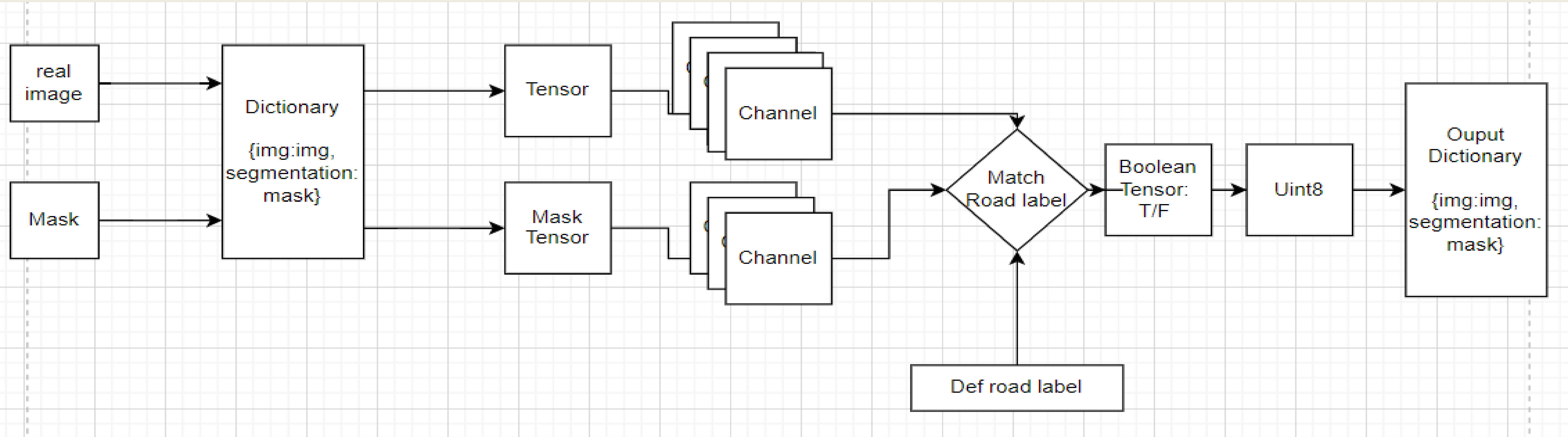
100 Epoch

After applying preprocessing steps such as resizing, label encoding, and batching, the data is visualized using Matplotlib.

Data Preprocessing



Data Preprocessing



Deep learning architectures for semantic segmentation

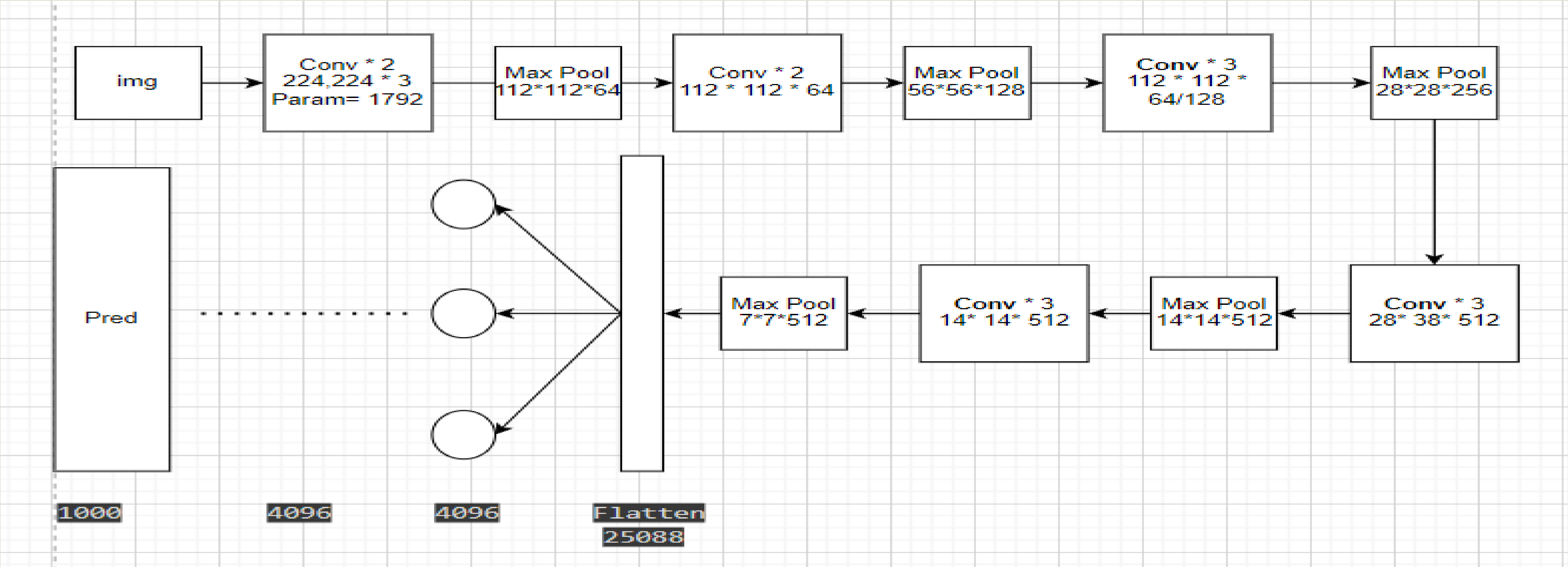
These architectures have been widely adopted and have demonstrated excellent performance in various semantic segmentation tasks across different domains, such as autonomous driving, medical imaging, and remote sensing.

1. **FULLY CONVOLUTIONAL NETWORKS (FCN)**
2. **U-NET**
3. **SEGNET**
4. **DENSEASPP**

MODELS

Deep learning architectures that use convolutional layers to automatically learn hierarchical features from input data. They excel in processing grid-like data, such as images, by applying learnable filters to capture local patterns and spatial relationships, enabling tasks like classification, segmentation, and object detection.

VGG Network



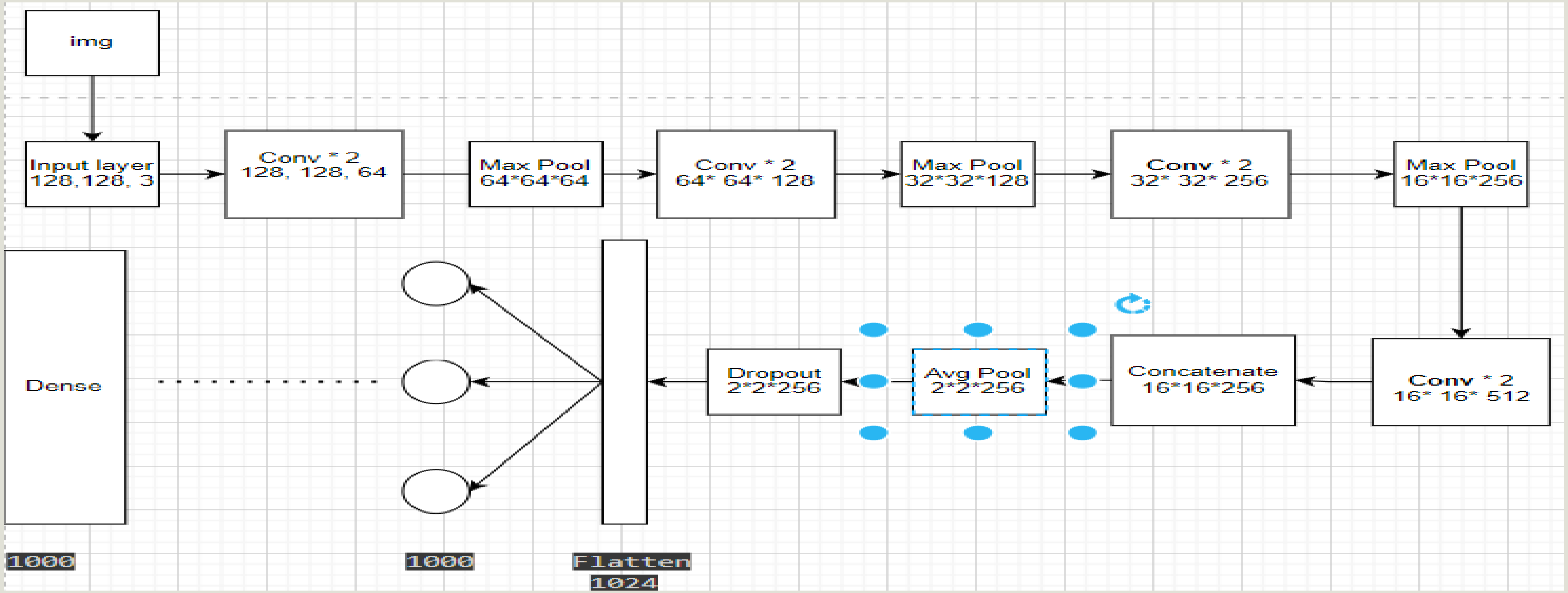
Overview

VGG (Visual Geometry Group) is a deep CNN architecture known for its simplicity and depth. It consists of stacked convolutional layers with small filters, followed by max-pooling and fully connected layers. VGG-16 and VGG-19 are popular versions. Despite newer architectures, VGG remains widely used for its effectiveness as a feature extractor.

Key Features

- Depth and Simplicity
- Effective Receptive Field
- Transfer Learning
- Robustness and Generalization
- Computational Considerations

GoogLeNet

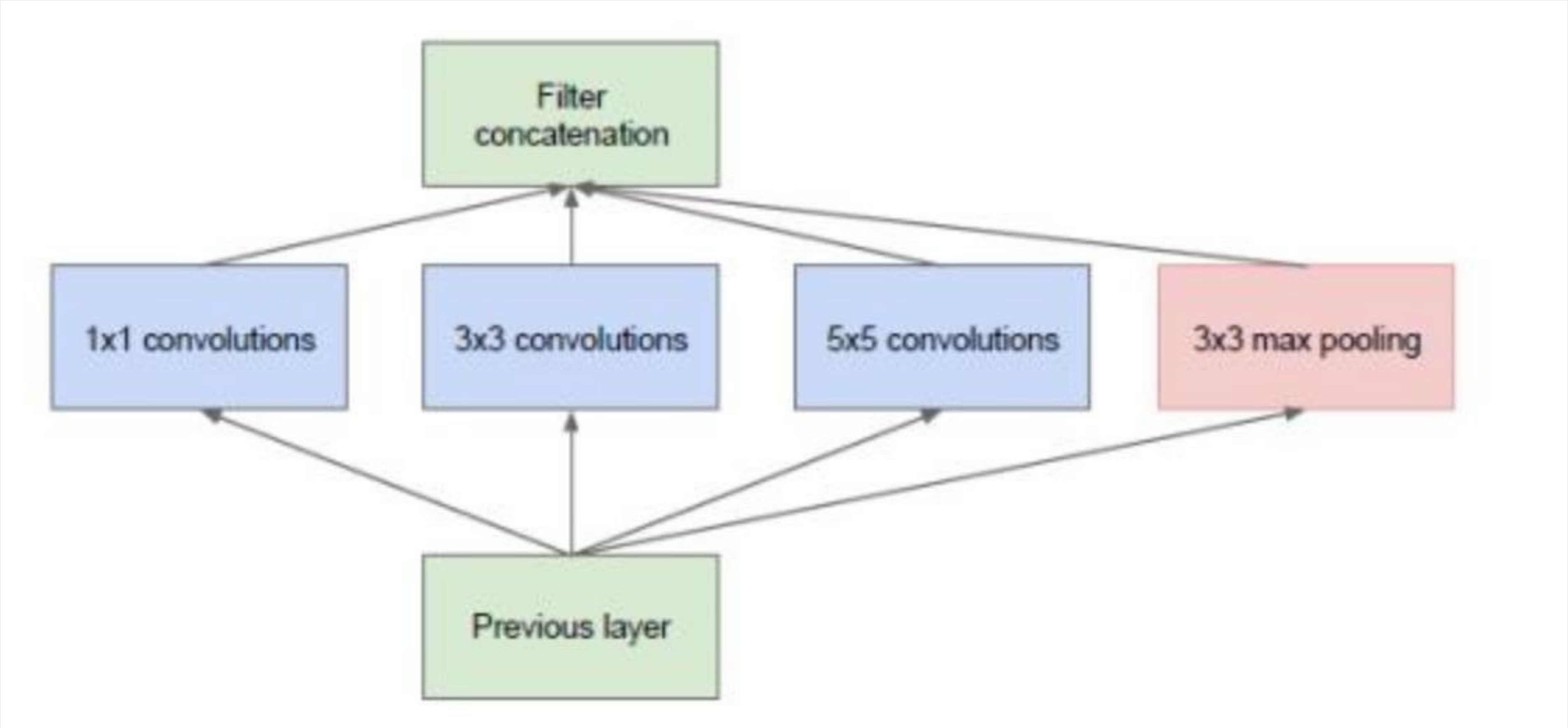


Overview

GoogLeNet, also known as Inception, is a deep CNN architecture developed by Google. It introduces the concept of Inception modules, which concatenate multiple convolutional filters of different sizes to capture multi-scale features. GoogLeNet is known for its efficiency and has significantly fewer parameters compared to other deep networks.

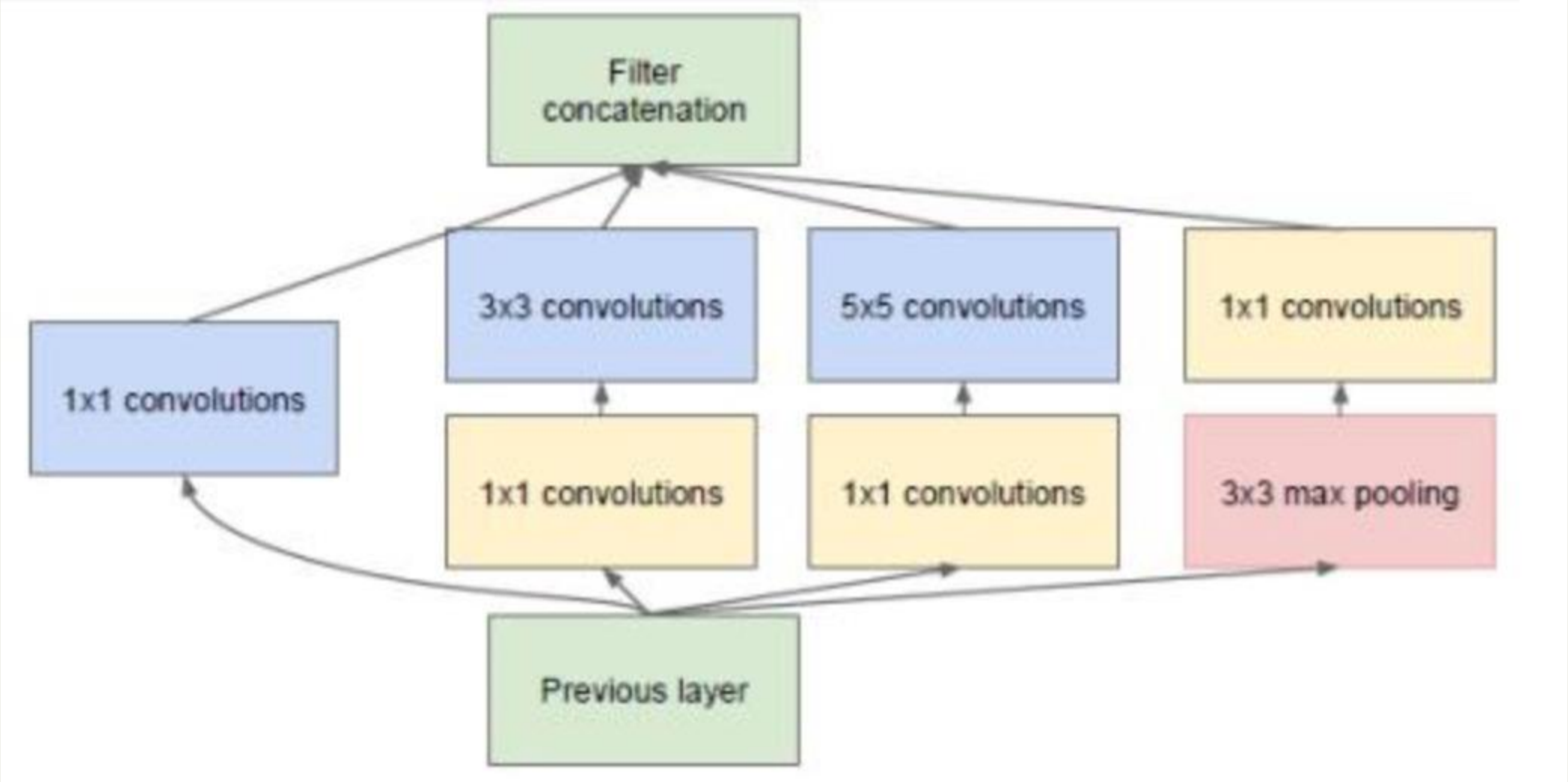
Key Features

- Inception Modules
- Dimensionality Reduction
- Multi-Scale Feature Extraction
- Auxiliary Classifiers
- Computational Efficiency
- Depth and Accuracy

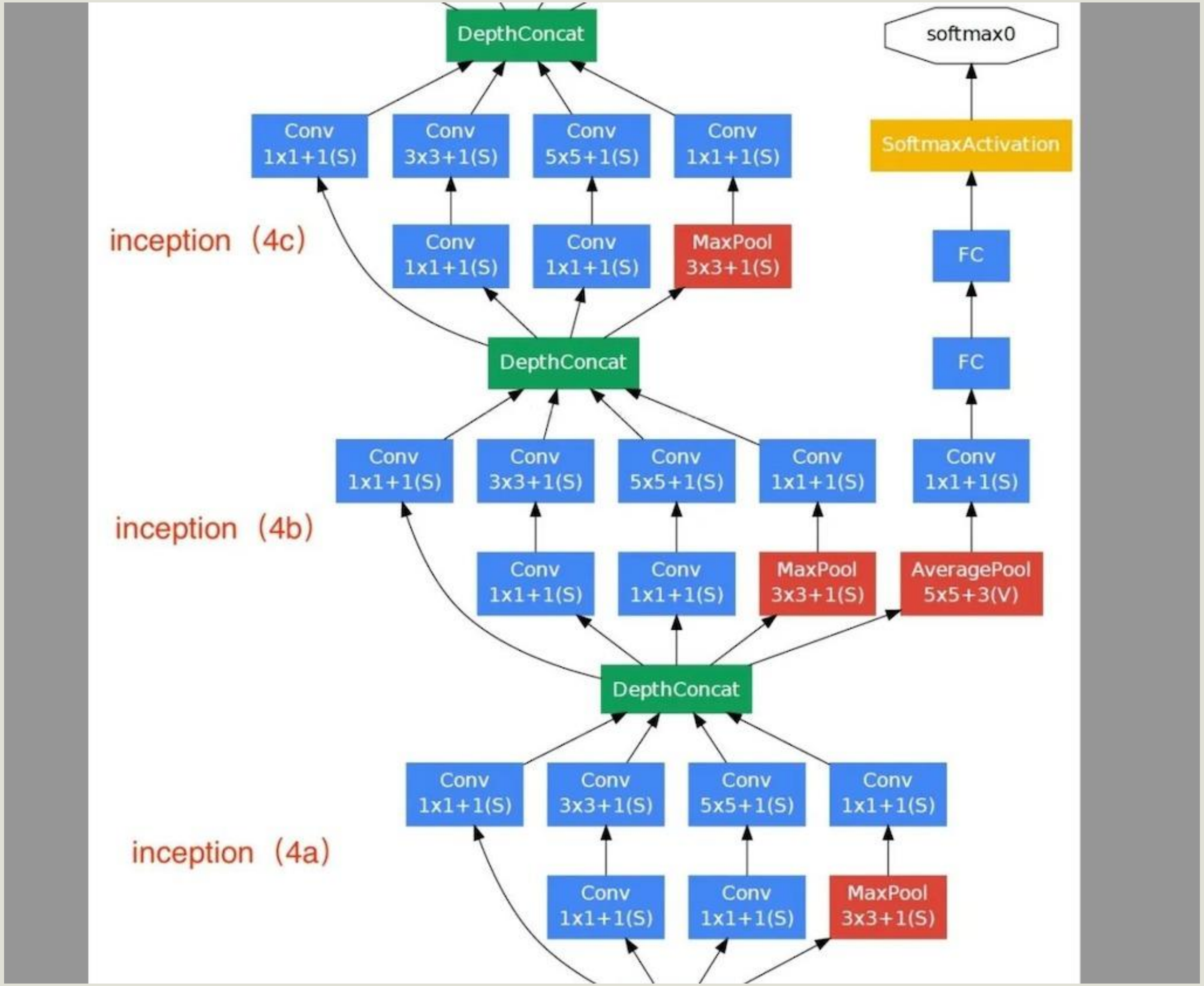


• Naïve Inception

• Inception V1

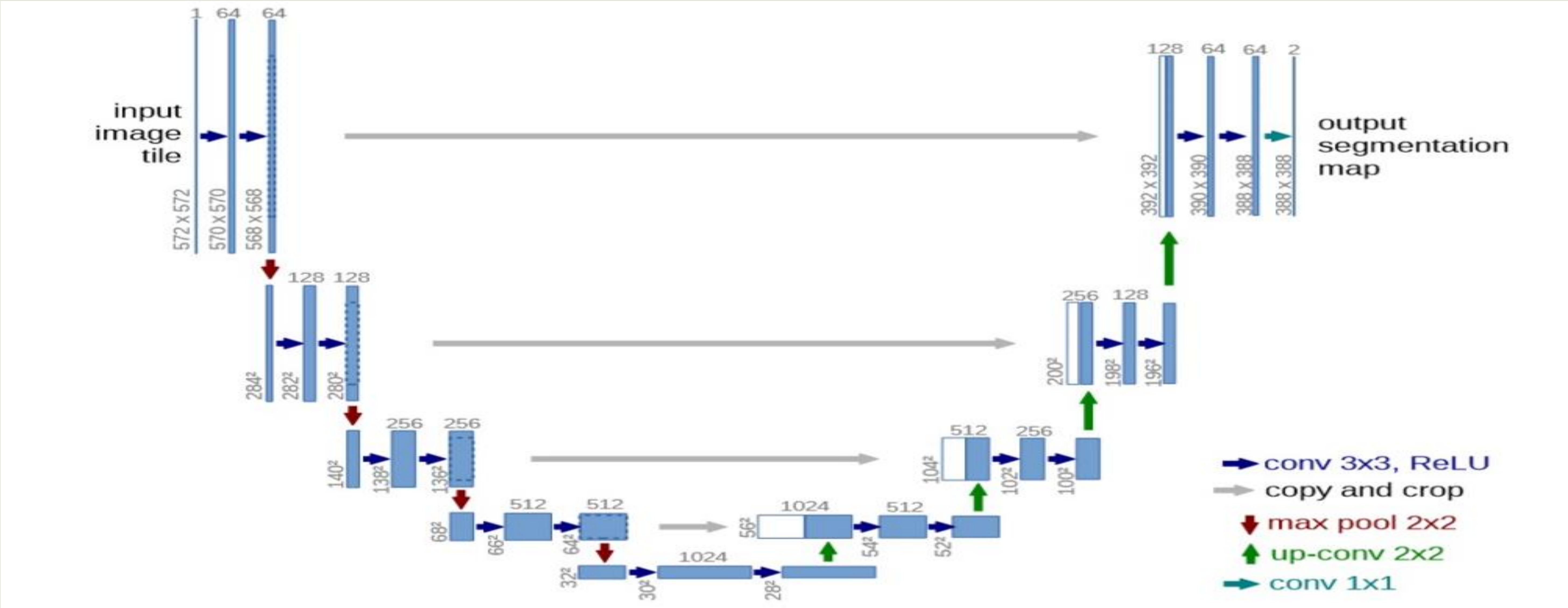


GoogLeNet



DepthConcat
softmax0
Conv
1x1+1(S)
Conv
3x3+1(S)
Conv
5x5+1(S)
Conv
1x1+1(S)
SoftmaxActivation
inception（4c）
Conv
1x1+1(S)
Conv
1x1+1(S)
MaxPool
3x3+1(S)
FC
FC
DepthConcat
Conv
1x1+1(S)
Conv
3x3+1(S)
Conv
5x5+1(S)
Conv
1x1+1(S)
Conv
1x1+1(S)
inception（4b）
Conv
1x1+1(S)
Conv
1x1+1(S)
MaxPool
3x3+1(S)
AveragePool
5x5+3(V)
DepthConcat
Conv
1x1+1(S)
Conv
3x3+1(S)
Conv
5x5+1(S)
Conv
1x1+1(S)
inception（4a）
Conv
1x1+1(S)
Conv
1x1+1(S)
MaxPool
3x3+1(S)

U-Net

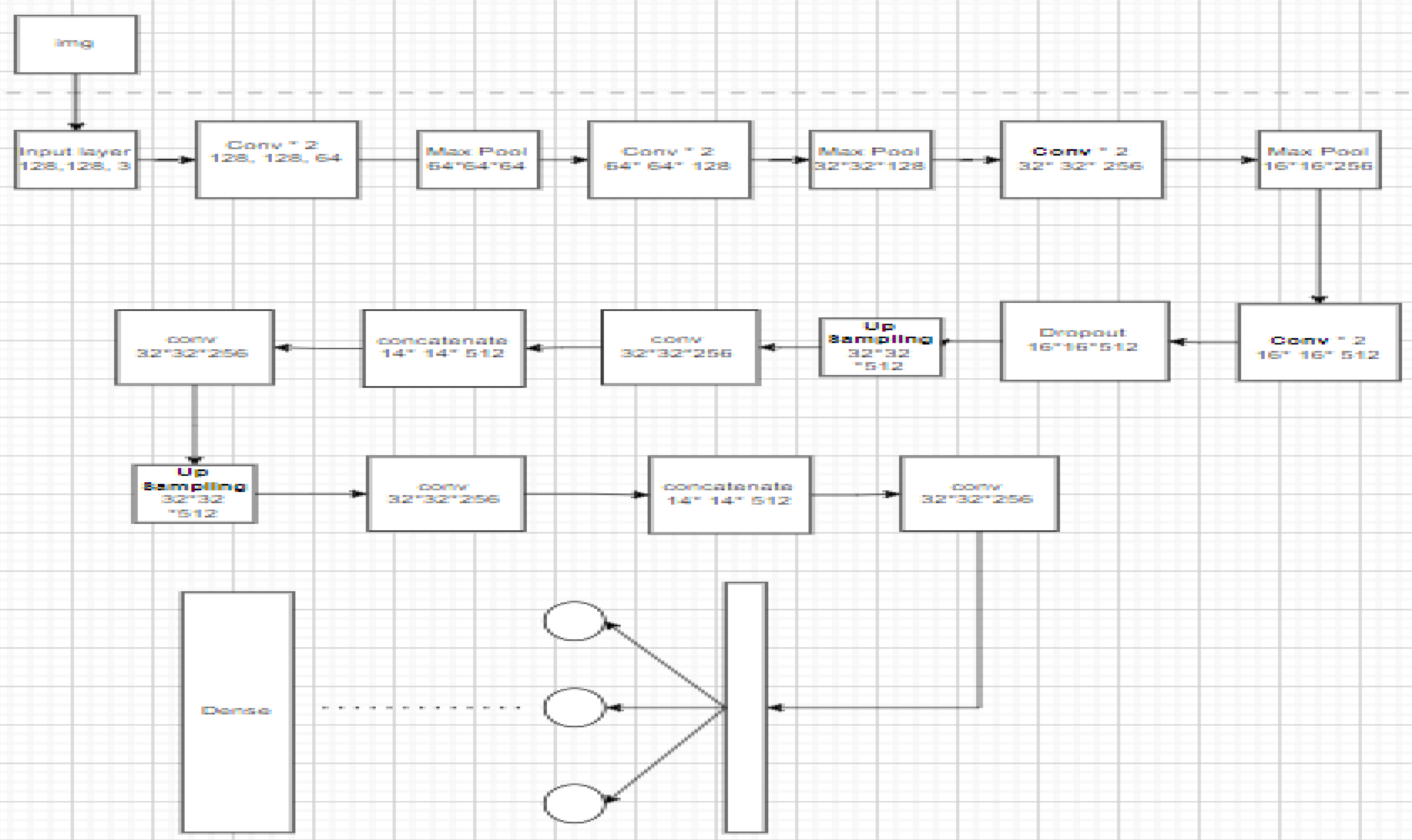


Overview

U-Net is a convolutional neural network architecture designed for biomedical image segmentation. It consists of an encoder-decoder structure with skip connections between corresponding layers. The encoder captures context, while the decoder enables precise localization. U-Net's symmetric architecture and skip connections allow for effective information propagation, making it popular for various segmentation tasks.

Key Features

- Encoder-Decoder Architecture
- Skip Connections
- Symmetric Structure
- Upsampling Techniques
- Effective for Limited Data
- Biomedical Image Segmentation
- Extensibility



img
Input layer
128,128, 3
Conv * 2
128, 128, 64
Max Pool
64*64*64
Conv * 2
64* 64* 128
Max Pool
32*32*128
Conv * 2
32* 32* 256
Max Pool
16*16*256
Conv * 2
16* 16* 512
Dropout
16*16*512
Up
Sampling
32*32
*512
conv
32*32*256
concatenate
14* 14* 512
conv
32*32*256
Up
Sampling
32*32
*512
conv
32*32*256
concatenate
14* 14* 512
conv
32*32*256
Dense

Unet + Inception

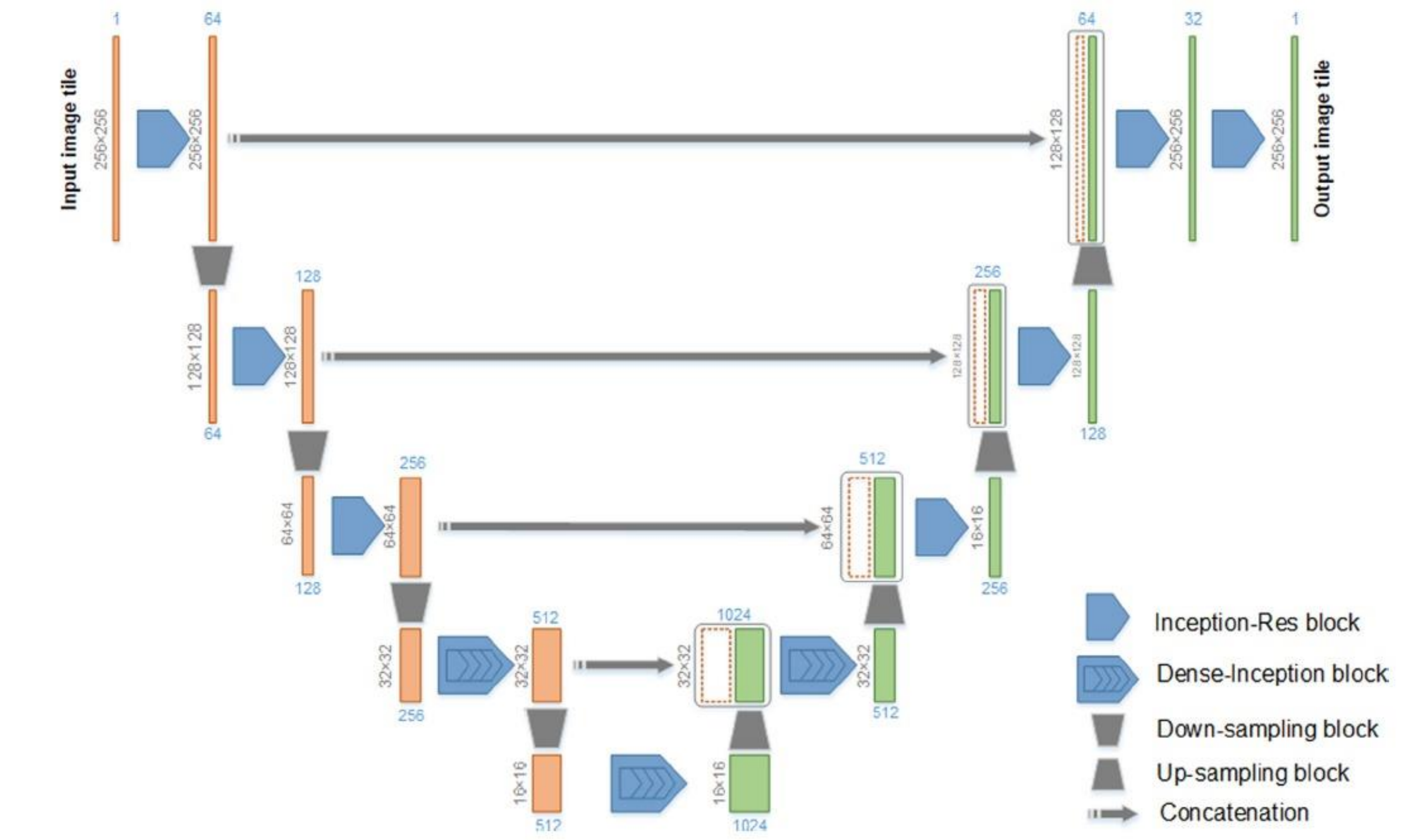


Fig. 1. Overall architecture of the DIU-Net model.

Experimental Setups & Results

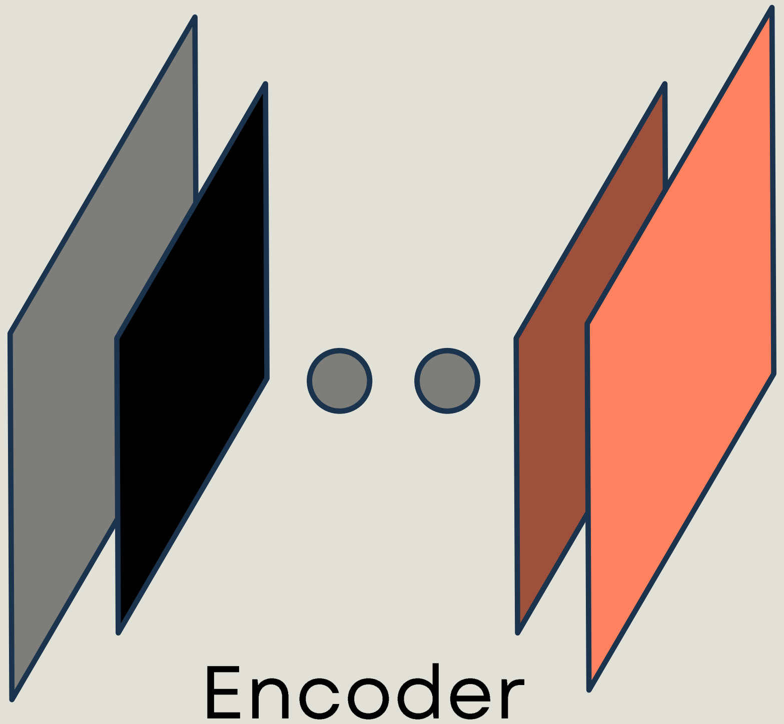
And popular architectures

Training Config.

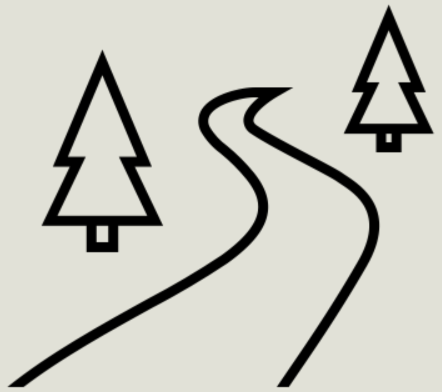
Configurations & Hyperparameters

VGG16 is chosen as the base model for road segmentation:

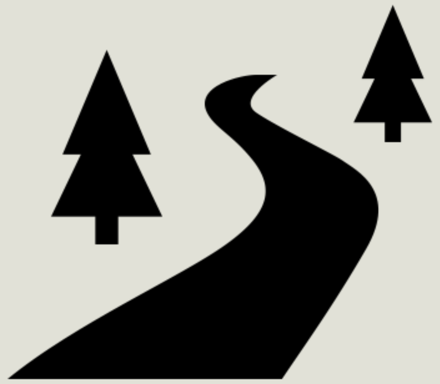
Input



Encoder
&
Decoder Layers



Output



Up-Sampling

Input Shape
(224,224,3)

Convolutional layers serve as encoder layers.
After encoding the input image into a set of feature maps, the decoder layers reconstruct the segmentation mask from these features

13 Convolution Layers

1000 Output Shape

0 Non-trainable params

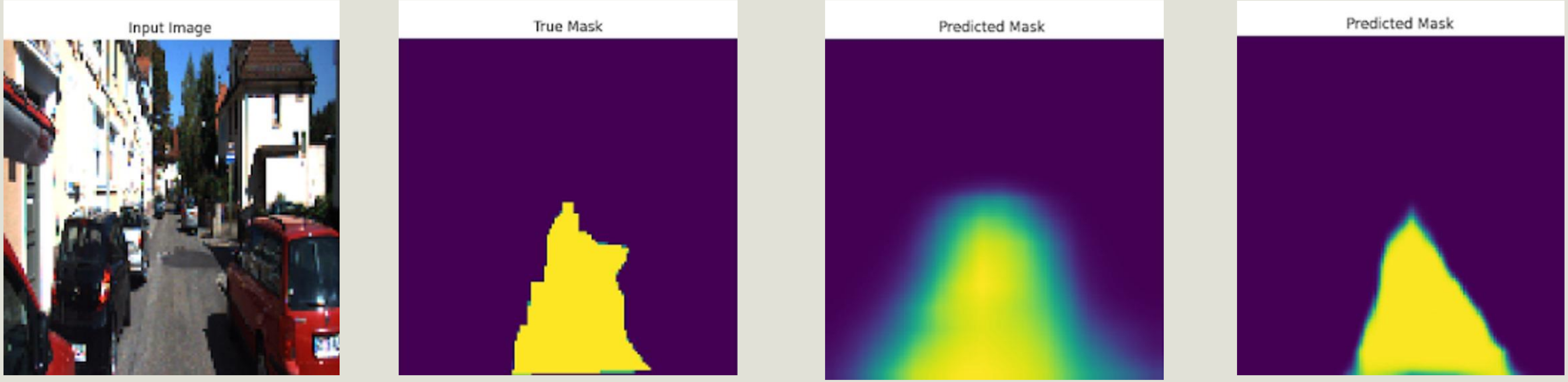
Total Parameters
138,357,544

Quantitative evaluation metrics

	VGG	GoogLeNet	U-Net
Accuracy Rate	99.95%	99.88	99.90
IOUs	0.046	0.4112	0.415

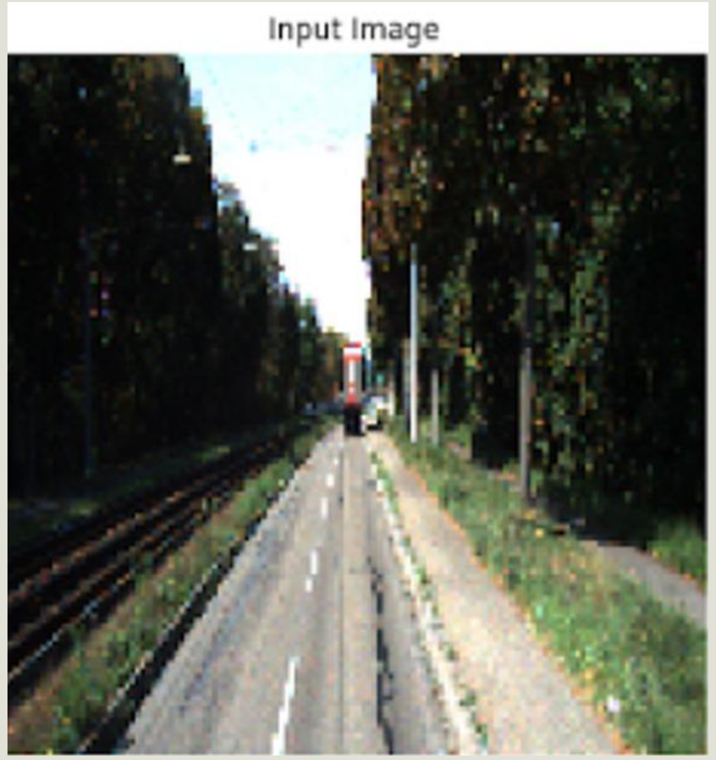
Qualitative Results

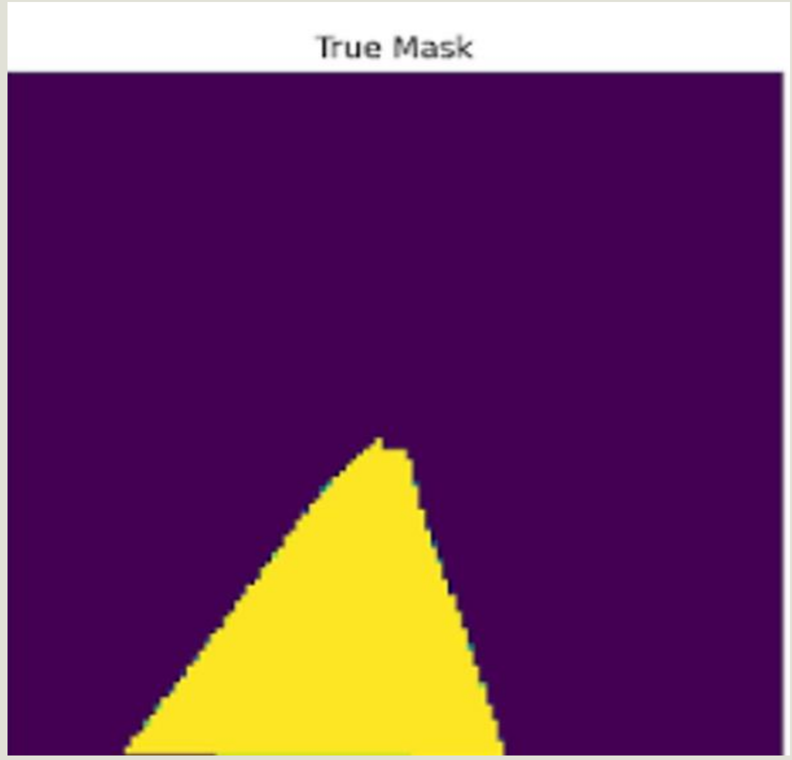
VGG

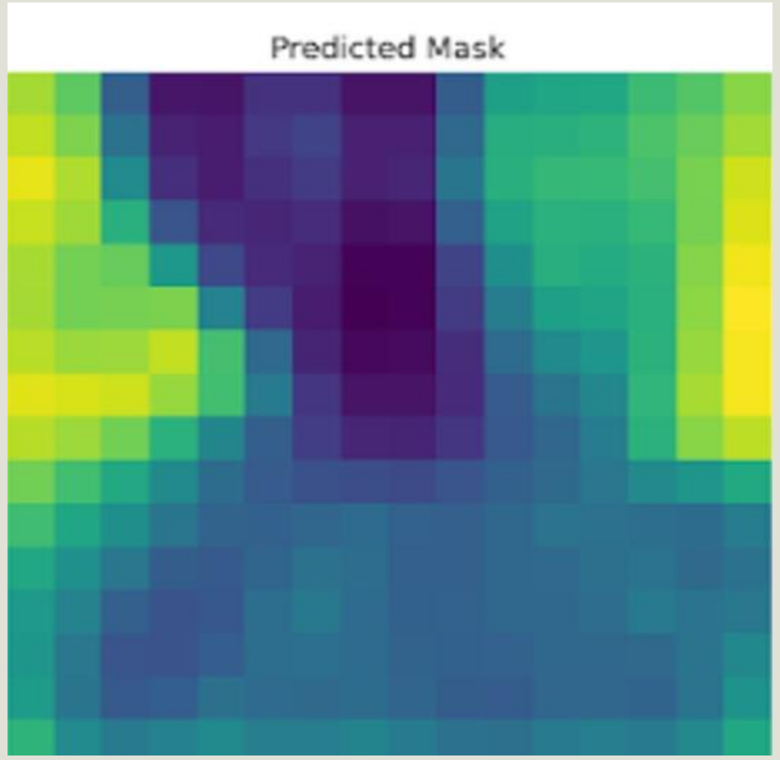


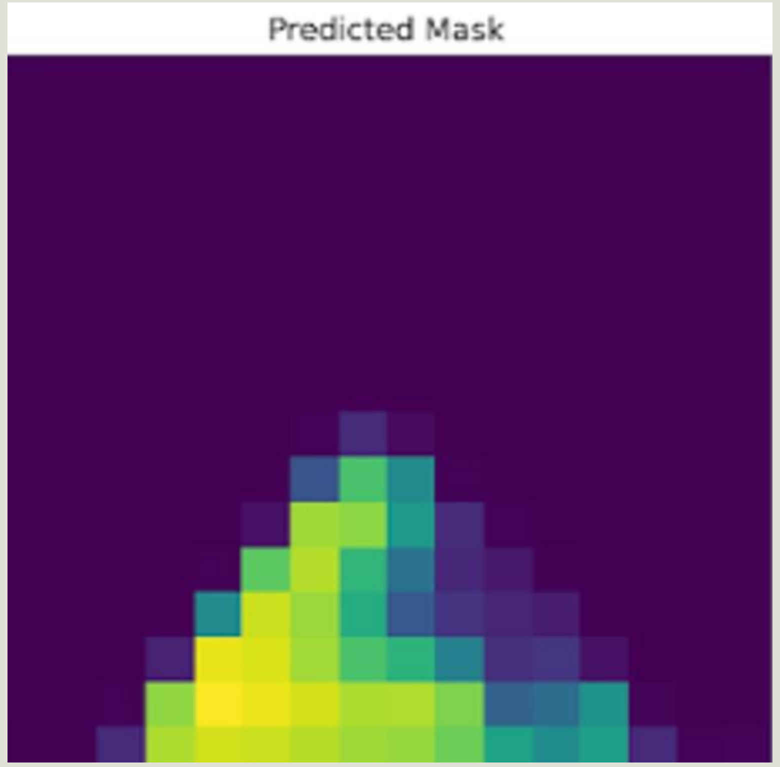
Qualitative Results

GoogLeNet



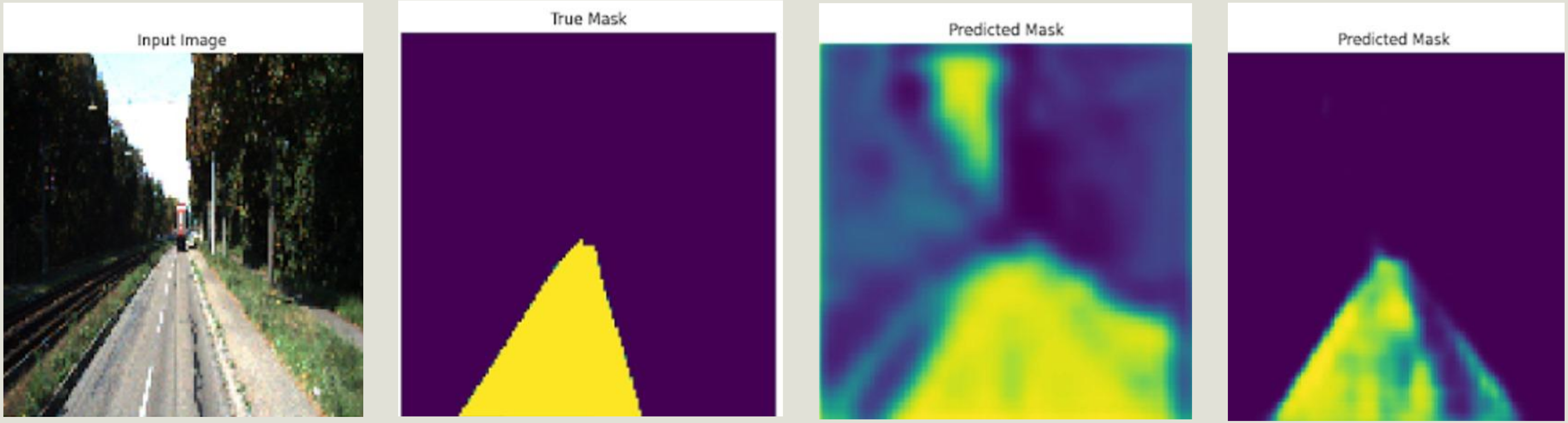






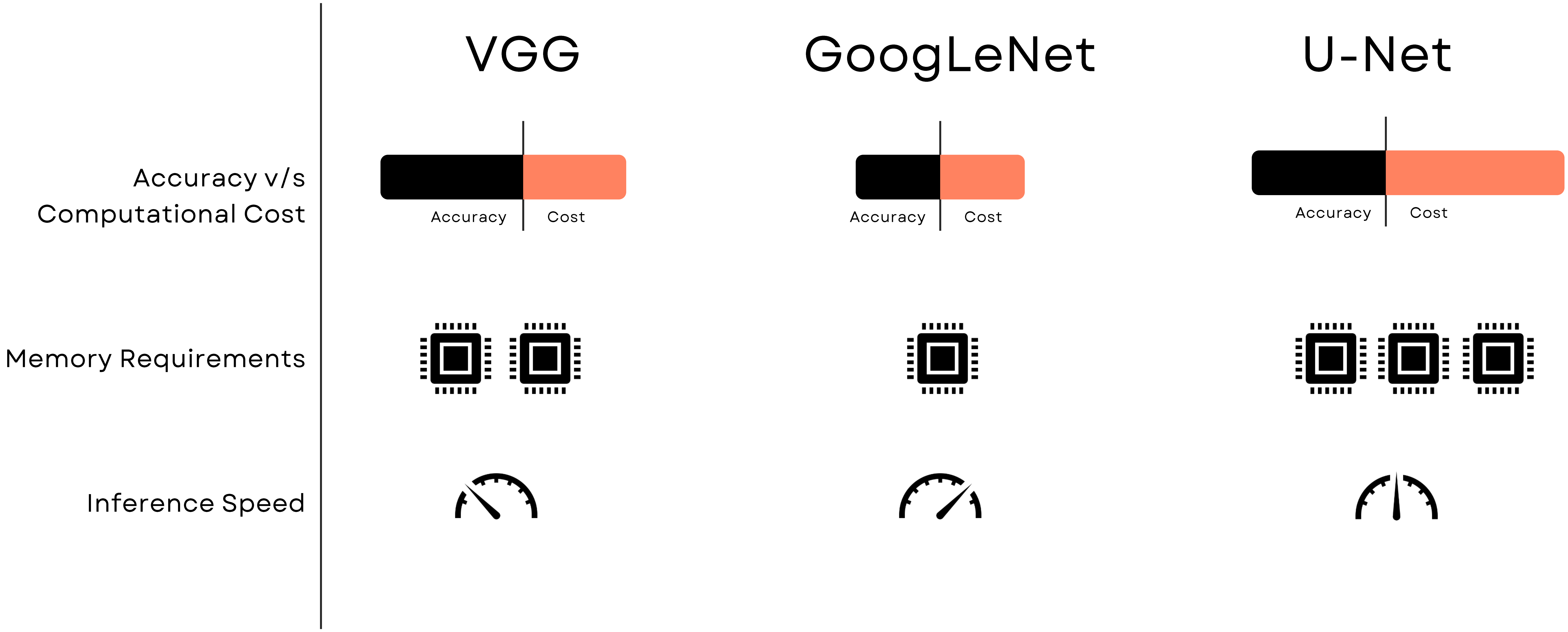
Qualitative Results

U-Net



How do these model stack against each other.?

Performance Trade-offs and Considerations



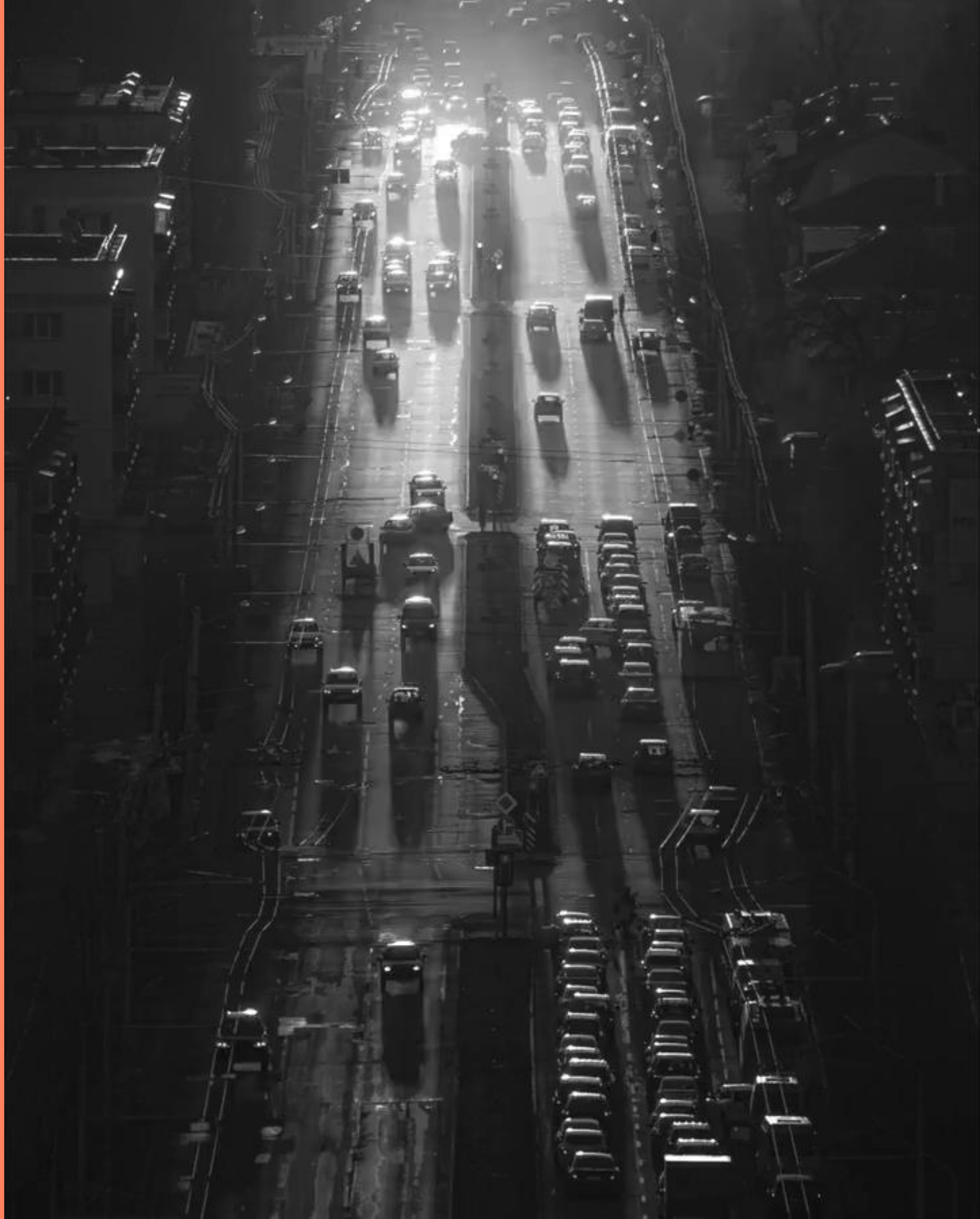
Suitability for Different Road Conditions and Scenarios

	VGG	GoogLeNet	U-Net
Urban Environments	Suitable for relatively structured roads with clear boundaries	Suitable for multi-scale features with varying object sizes	Fine-grained details, identify irregular shapes, pedestrians and vehicles
Highway and Rural Roads	Performs well in highways and rural roads with homogeneous details	Captures highway and rural roads with varying road widths	Suitable especially when accurate segmentation of road boundaries and obstacles is crucial
Challenging Lighting Conditions	Performs well in bright light. May struggle in low-light or varying illumination	Captures details in different scales, making it more robust for challenging light conditions	Preserves spatial information, making it resilient to light variations

Conclusion

Recap of the key findings and insights:

Since we used the custom VGG model, the highest accuracy is 99.95%, followed by U-Net with an accuracy 99.90%, and GoogLeNet with 99.88%.



THE TEAM

Nidhi Jha

RUID:

Advait Kalsekar

RUID:

Vishal Pawar

RUID

Parth Thaker

RUID

Davina

RUID

Thank you