Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment

Assignment 2 Presentation

Alireza Imani
M.Sc. Student in Electrical and Computer Engineering
Schulich School of Engineering

2021-10-21

UNIVERSITY OF CALGARY

Semantic Segmentation

Karunakaran, D. (2018). Semantic segmentation — Udaity's self-driving car engineer nanodegree [Image]. Retrieved 20 October 2021, from https://medium.com/intro-to-artificial-intelligence/semantic-segmentation-udaitys-self-driving-car-engineer-nanodegree-c01eb6eaf9d.

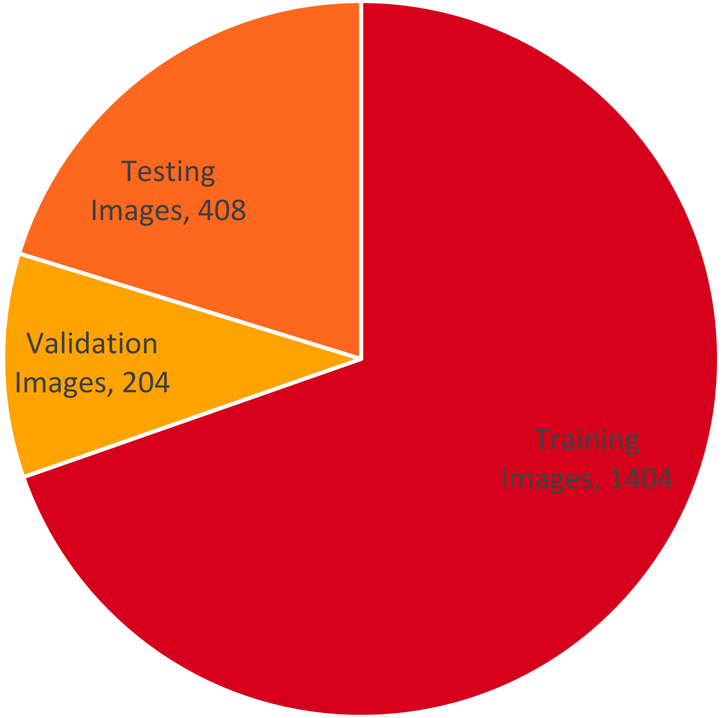
- The process of assigning pre-defined label or class to each pixel of an image which is also known as pixel level classification.

Semantic scene segmentation of unstructured driving environment

- Ambiguous road boundaries
- Unmarked or incompletely delineated lanes
- Wear and tear of road infrastructure
- High within class diversity
- Less adherence to traffic rules, etc.

Dataset: IDD Lite

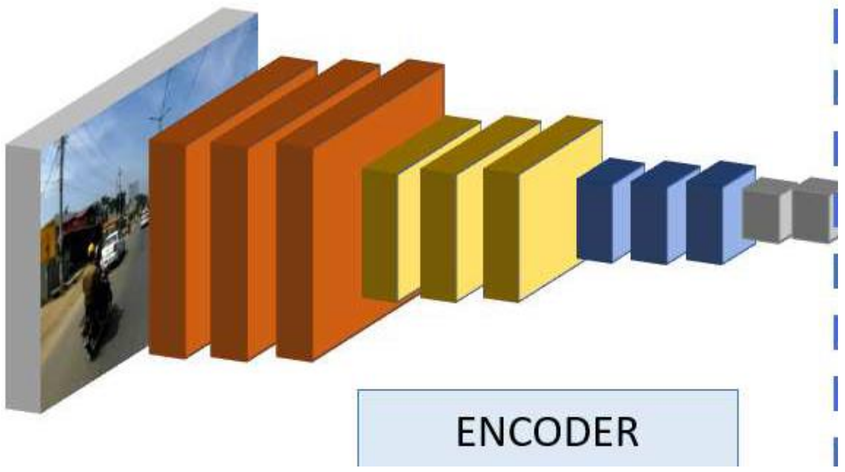
- A subsampled version of India Driving Dataset (IDD) for resource constrained training like lack of memory resources and high-end GPU.
- Similar label statistics as IDD and less number of labels
- Less than 50MB in size
- Resolution of 320×227
- 7 classes: Drivable, Non-drivable, Living things, Vehicles, Road-Side Objects, Far objects and Sky.



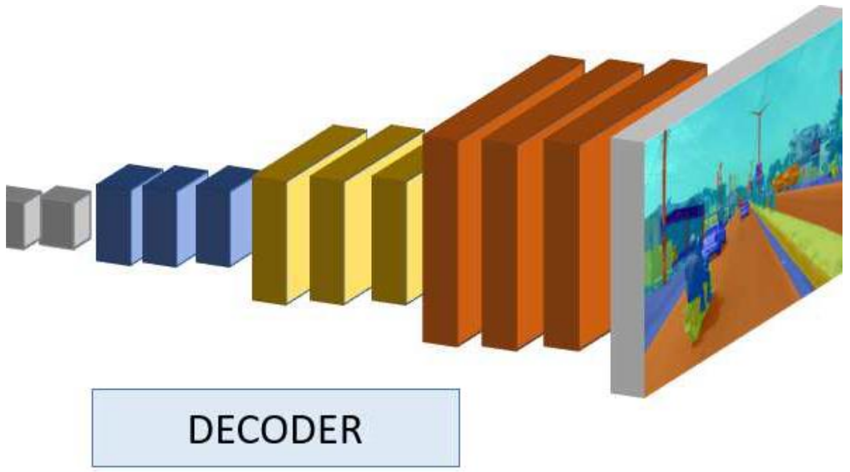
Challenges

- Ambiguous road boundaries that have muddy terrain, which is also drivable
- The diversity of vehicles and pedestrians and their locations on images
- No strict adherence to traffic rules
- Lightening condition
  - Mid-day
  - Dawn
  - Dusk

Encoder-Decoder Architecture



- A **CNN** which extracts the features from original image by progressively down-sampling it.
  - ResNet
  - MobileNet
  - InceptionResNetV2



- A set of layers that up-samples the feature map of encoder to recover spatial information

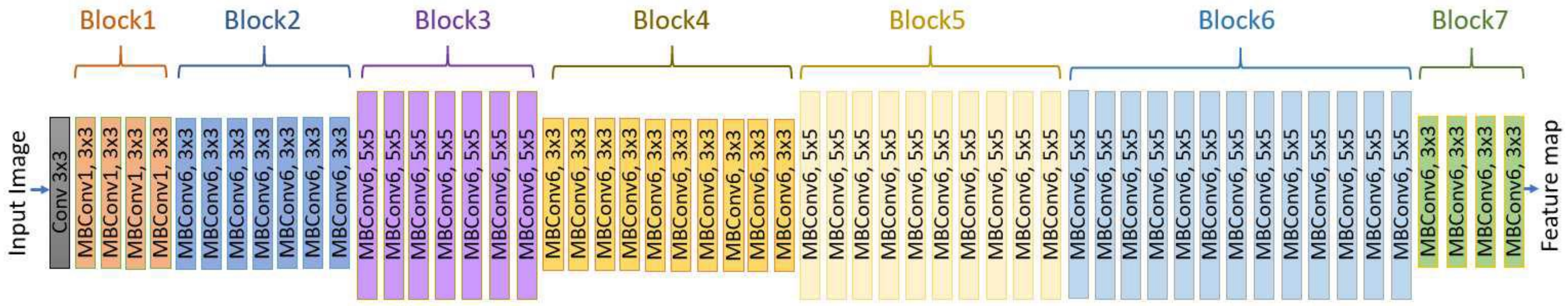
EfficientNet

- CNN architectures development depends on the available resources
- Scaling occurs to achieve improved performance
- Traditional Practice: Increasing the CNN width, depth or the input image resolution arbitrarily
- EfficientNet: A Novel compound scaling method which uniformly scales the network depth, width and resolution for improved performance with a fixed set of scaling

EfficientNet Cont.

UNIVERSITY OF CALGARY

- EfficientNetB0 to EfficientNetB7
- Basic building block: Mobile Inverted Bottleneck Convolution (MBConv)



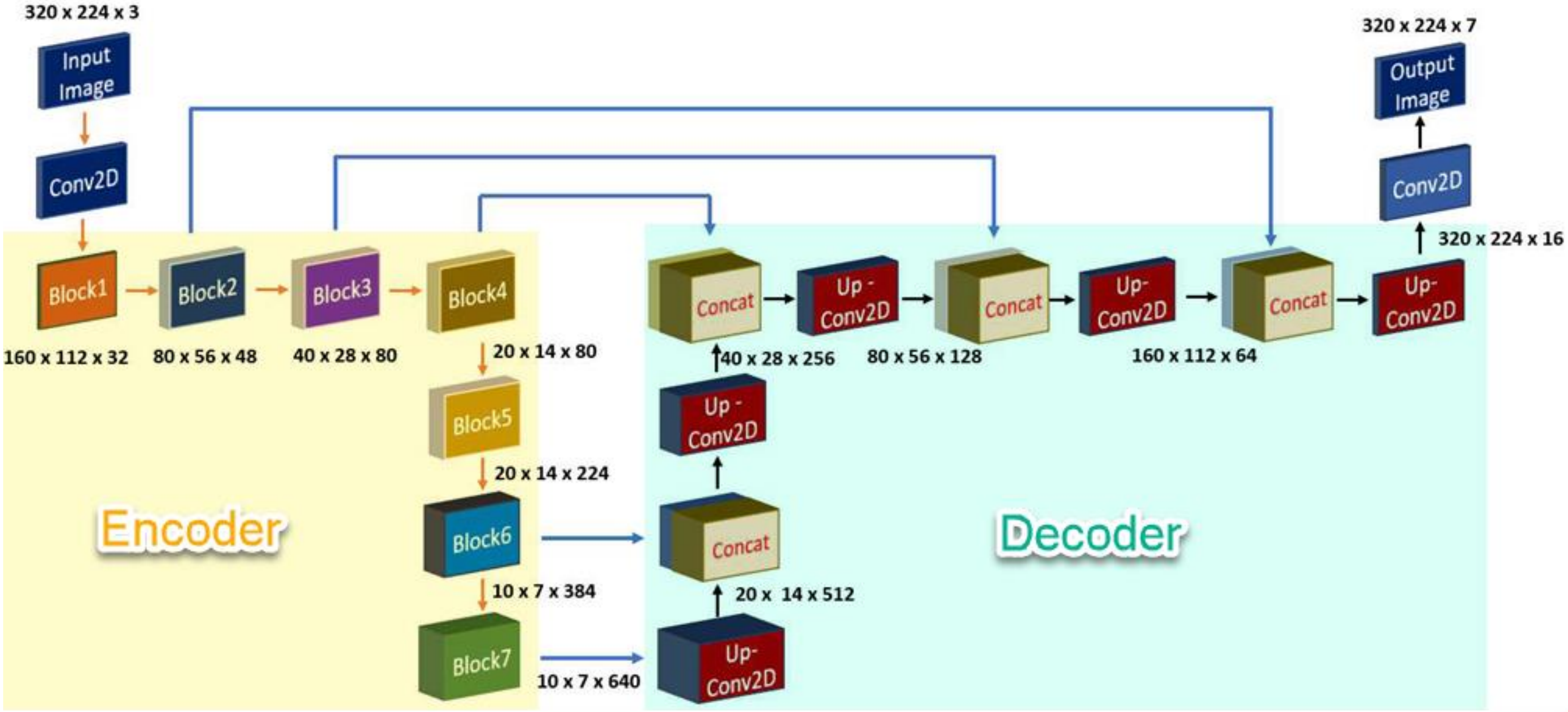
Architecture of **EfficientNetB7** with MBConv as basic building blocks
MBConv**6** uses RELU**6** activation function
MBConv uses RELU activation function
8.4× **smaller** and 6.1× **faster** than the best existing CNN

UNet Decoder

UNIVERSITY OF CALGARY

- A symmetric **U shaped** fully convolutional neural network originally developed for biomedical image segmentation
- Has two paths:
    - Contraction path (encoder): a stack of convolution, activation and pooling layers to capture the context in the input image.
    - Expansion path (decoder): combines the high-level features and spatial information by a sequence of up-convolutions and concatenation with corresponding feature maps from the contracting path.
- Eff-UNet: EfficientNet as an **encoder** in contracting path instead of conventional set of convolution layers.

Eff-UNet Architecture



Implementation & Training

UNIVERSITY OF CALGARY

- Tensorflow 2.0
- Resized images to 320×224 (originally 320x227)
- Batch size of 4 - 10 epochs
- The pre-trained weights of **EfficientNetB7** on **ImageNet** are used for initialization in encoder (Transfer learning!)
- Augmented the dataset with various transformations like brightness, contrast, saturation, shear, etc.
- Using **ADAM optimizer** with a learning rate of 0.0001
- **(Jaccard + binary cross-entropy)** is used as loss function

Results

- Evaluated in terms of mean Intersection over Union (mIoU) also called as Jaccard index:
- $IoU = \frac{TP}{TP+FP+FN}$

Network Architecture	Validation mIoU	Test mIoU
Dilated ResNet18 [33]	0.5503	-
ERFNet [19]	0.6614	-
DeepLabV3+ with ResNet18 Encoder	0.6304	0.5614
DeepLabV3+ with ResNet50 Encoder	0.6425	0.5733
UNet with ResNet34 Encoder	0.6781	0.6009
UNet with ResNet50 Encoder	0.6859	0.6076
UNet with InceptionResNetV2 Encoder	0.7247	0.6175
UNet with EfficientNetB5 Encoder	0.7072	0.6087
UNet with EfficientNetB7 Encoder	**0.7376**	**0.6276**

Decoders

Encoders

Baseline Performance

Segmentation Results

- First Col: Original Photo
- Second Col: Ground truth
- Third Col: Predicted segmentation map

Thank you for your attention!

All the contents and uncited raw images are referenced to the presented paper.

Alireza Imani
alireza.imani@ucalgary.ca