Module 2: Building Blocks for Image Recognition

Video 11 : EfficientNet + Hands on

Analytics Vidhya

EfficientNet

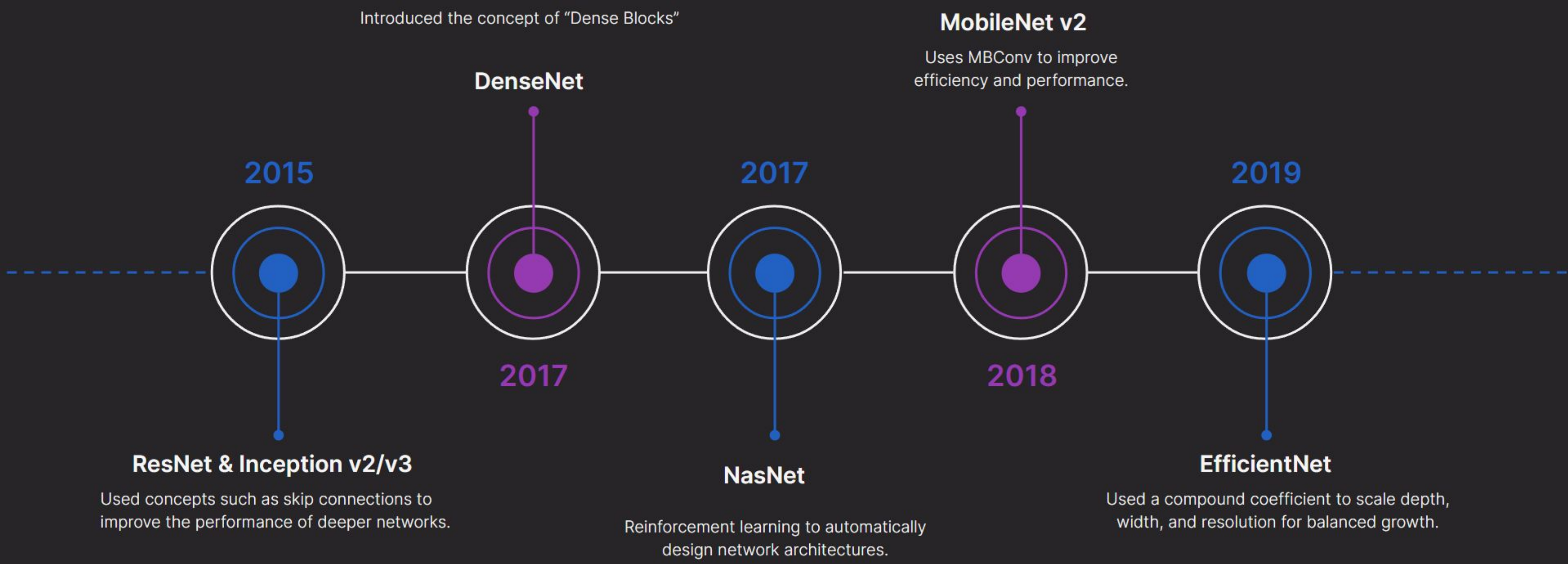
- Developed by **Mingxing Tan and Quoc V. Le** in **2019**.
- Introduced compound scaling & redefined how CNNs are scaled.

Mingxing Tan

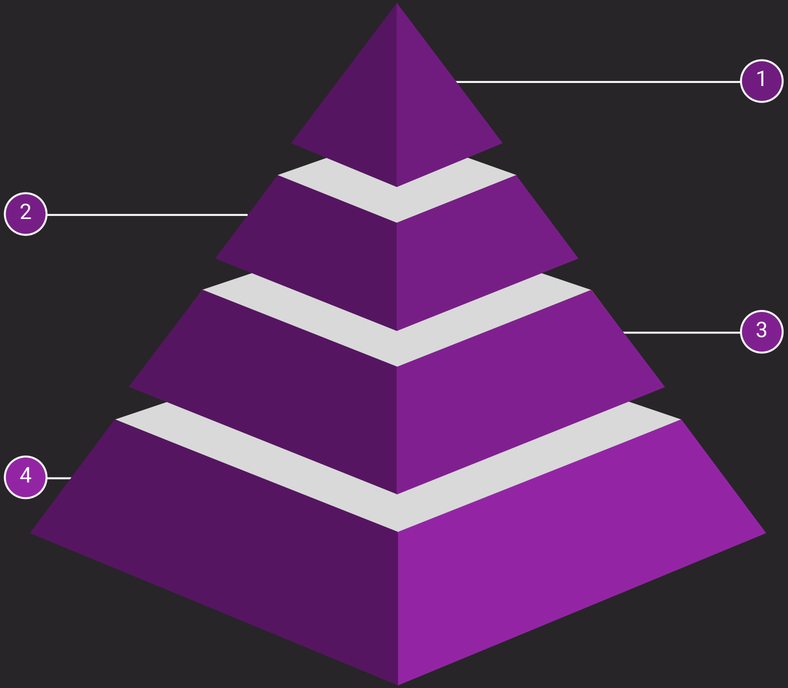
Quoc V. Le

EfficientNet vs DenseNet

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
DenseNet121	33	75.0%	92.3%	8.1M	242	77.1	5.4
DenseNet169	57	76.2%	93.2%	14.3M	338	96.4	6.3
DenseNet201	80	77.3%	93.6%	20.2M	402	127.2	6.7
EfficientNetB0	29	77.1%	93.3%	5.3M	132	46.0	4.9
EfficientNetB1	31	79.1%	94.4%	7.9M	186	60.2	5.6
EfficientNetB2	36	80.1%	94.9%	9.2M	186	80.8	6.5
EfficientNetB3	48	81.6%	95.7%	12.3M	210	140.0	8.8
EfficientNetB4	75	82.9%	96.4%	19.5M	258	308.3	15.1
EfficientNetB5	118	83.6%	96.7%	30.6M	312	579.2	25.3
EfficientNetB6	166	84.0%	96.8%	43.3M	360	958.1	40.4
EfficientNetB7	256	84.3%	97.0%	66.7M	438	1578.9	61.6


2015
ResNet & Inception v2/v3
Used concepts such as skip connections to improve the performance of deeper networks.
2017
DenseNet
Introduced the concept of "Dense Blocks"
2017
NasNet
Reinforcement learning to automatically design network architectures.
2018
MobileNet v2
Uses MBConv to improve efficiency and performance.
2019
EfficientNet
Used a compound coefficient to scale depth, width, and resolution for balanced growth.

Key Innovations in EfficientNet



Scaling
Applies Compound Scaling based on a set of fixed coefficients.

MBConv
Reduces parameters and enhances efficiency through depthwise convolutions.

Squeeze & Excitation
Recalibrates channel-wise features.

Swish Activation Function
Captures complex patterns and enhances performance.

1. Compound Scaling

- Depth (d) refers to the number of layers.
- Width(w) refers to the number of feature maps.
- Resolution(r) refers to the size of input images.
- Compound coefficient (ϕ) is a positive integer that determines the overall scaling.

1. Compound Scaling

EfficientNet B0 Values

$$\alpha = 1.2^{\phi}$$
$$\beta = 1.1^{\phi}$$
$$\gamma = 1.15^{\phi}$$

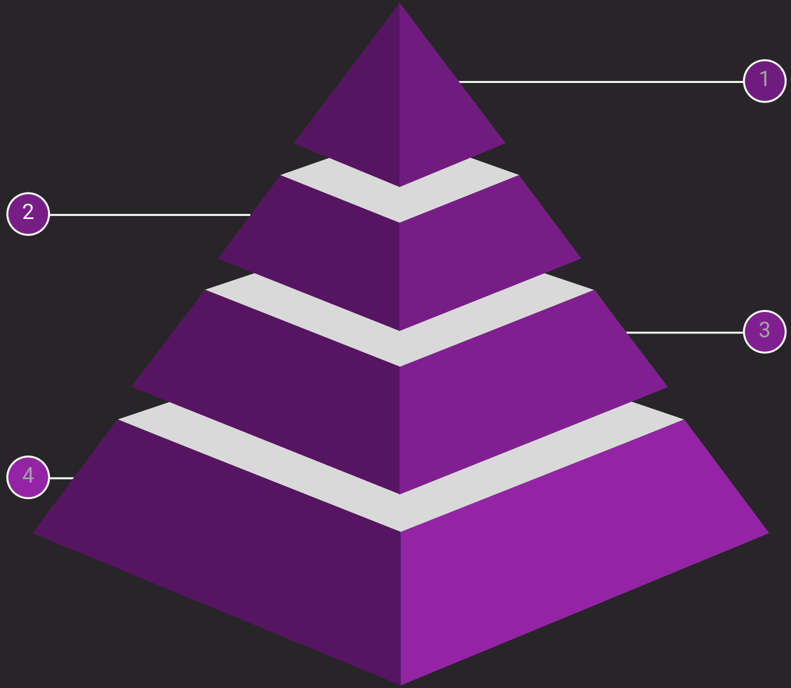
Resolution for EfficientNet B1 = $224 * \gamma^{\phi} = 224 * 1.15 = 257$

- $\phi=1$

Where:

- α is the scaling factor for depth.
- β is the scaling factor for width.
- γ is the scaling factor for resolution.
- ϕ is the compound coefficient, a positive integer that determines the overall scaling.

Key Innovations in EfficientNet



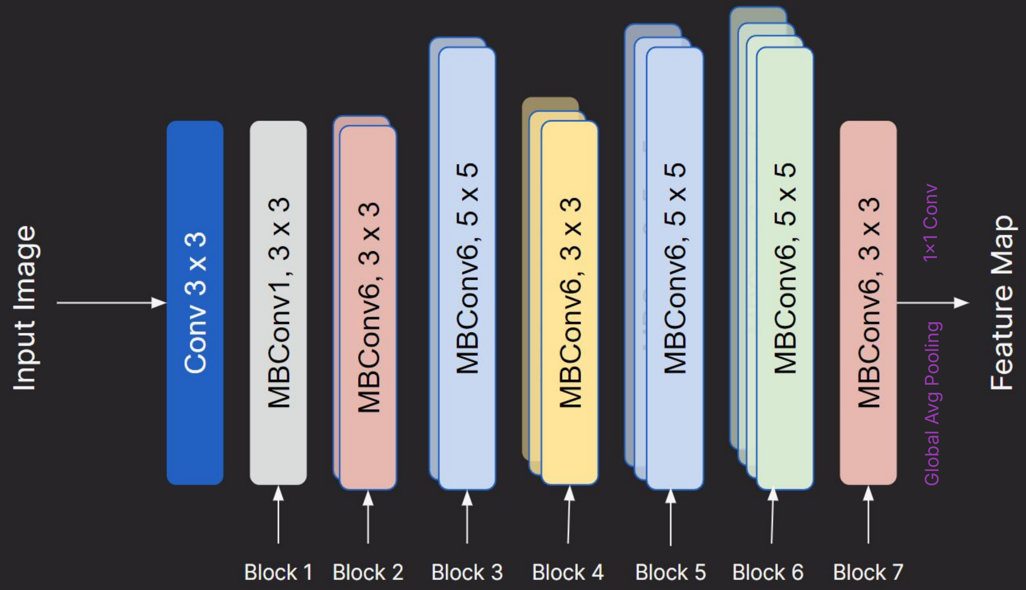
Scaling
Applies Compound Scaling based on a set of fixed coefficients.

MBConv
Reduces parameters and enhances efficiency through depthwise convolutions.

Squeeze & Excitation
Recalibrates channel-wise features.

Swish Activation Function
Captures complex patterns and enhances performance.

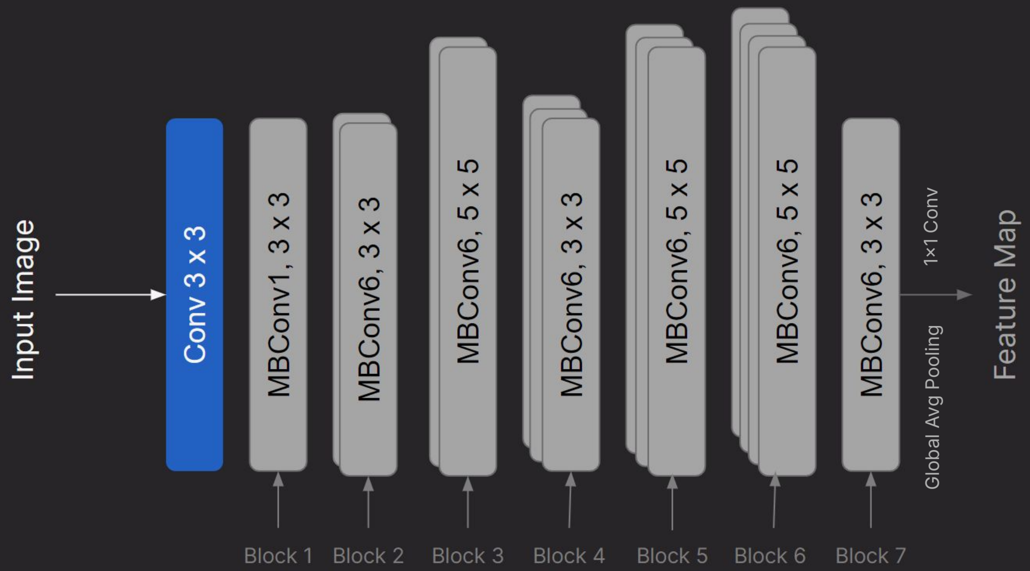
EfficientNet B0 Architecture



EfficientNet B0 Architecture

Initial Layer

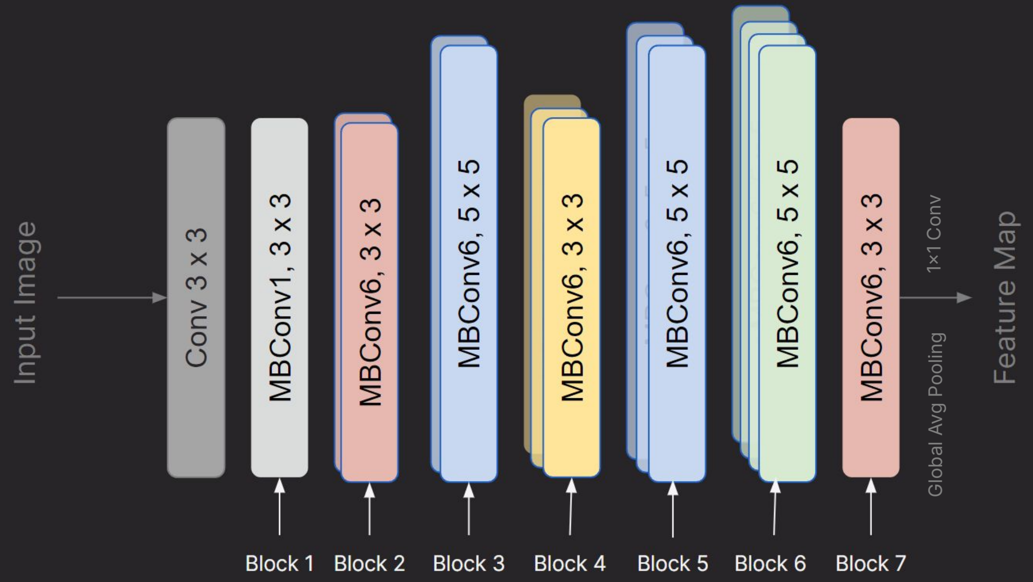
- 3×3 convolutional layer with 32 filters and a stride of 2.
- Batch Normalization and ReLU applied to stabilize learning and introduce non-linearity.



2. MBConv Blocks

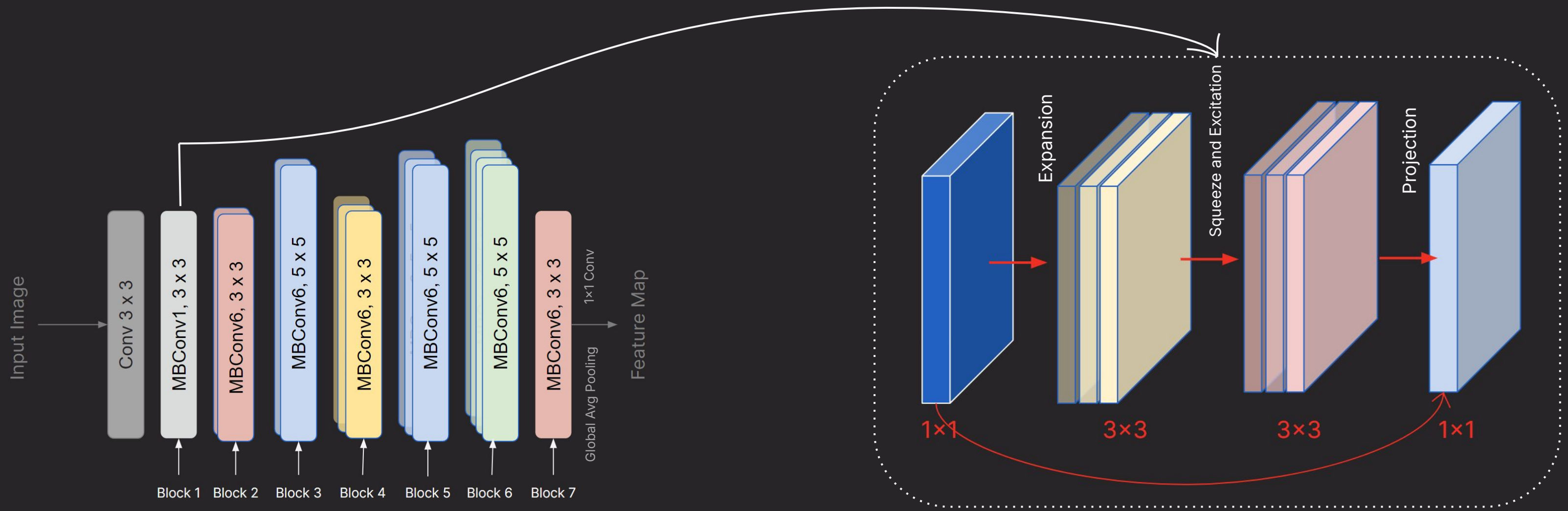
Mobile Inverted Bottleneck Convolution

- Builds upon MobileNetV2

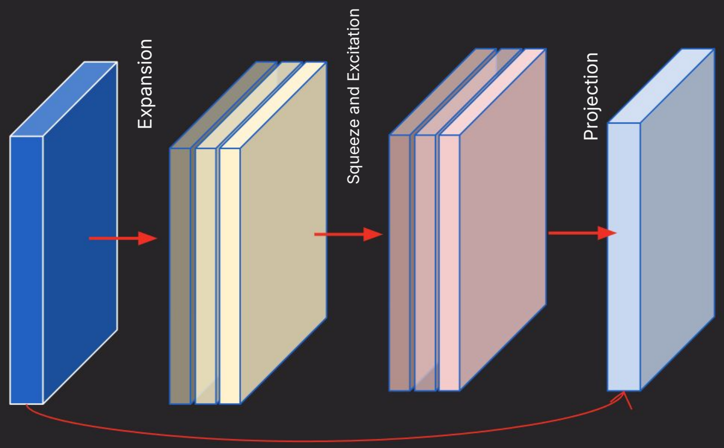


2. Mobile Inverted Bottleneck Convolution

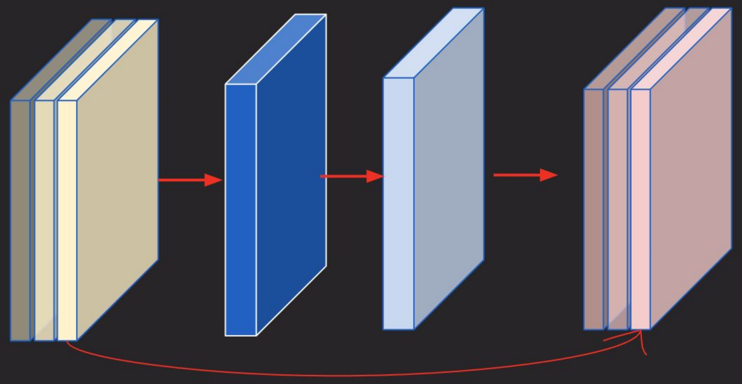
- **Expansion Phase:** Starts by expanding input feature maps using a 1×1 convolution
- **Depthwise Convolution:** Processes each channel separately; computationally efficient.
- **Projection Phase:** Reduces the number of channels back to the original size.



2. Mobile Inverted Bottleneck Convolution



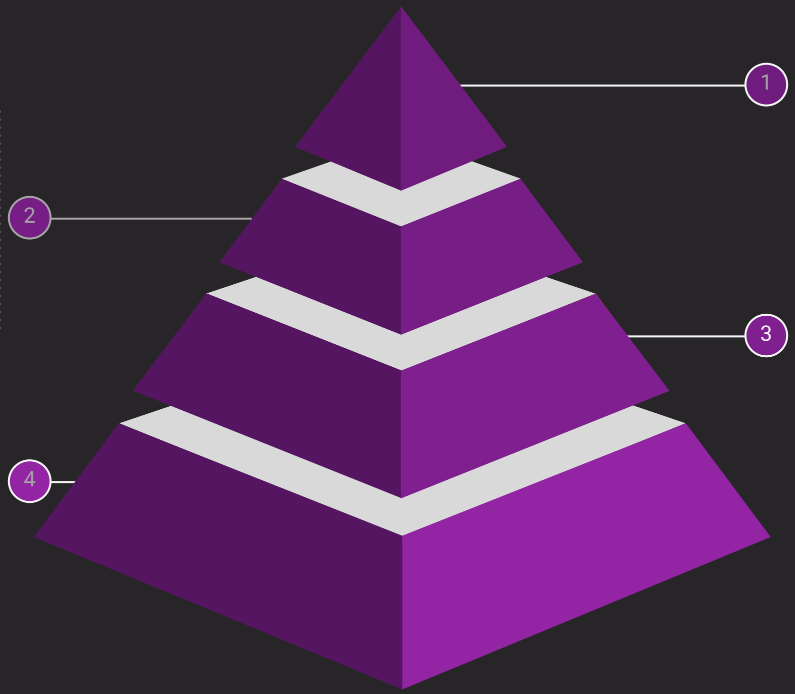
Efficient Net B0



ResNet

Analytics Vidhya

Key Innovations in EfficientNet



Scaling
Applies Compound Scaling based on a set of fixed coefficients.

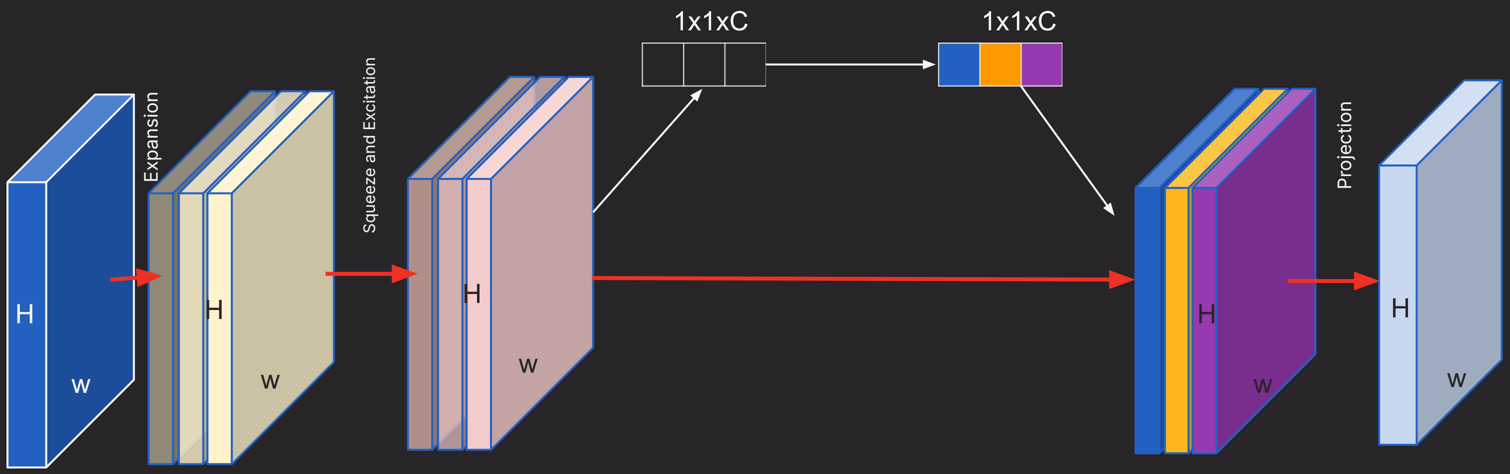
MBConv
Reduces parameters and enhances efficiency through depthwise convolutions.

Squeeze & Excitation
Recalibrates channel-wise features.

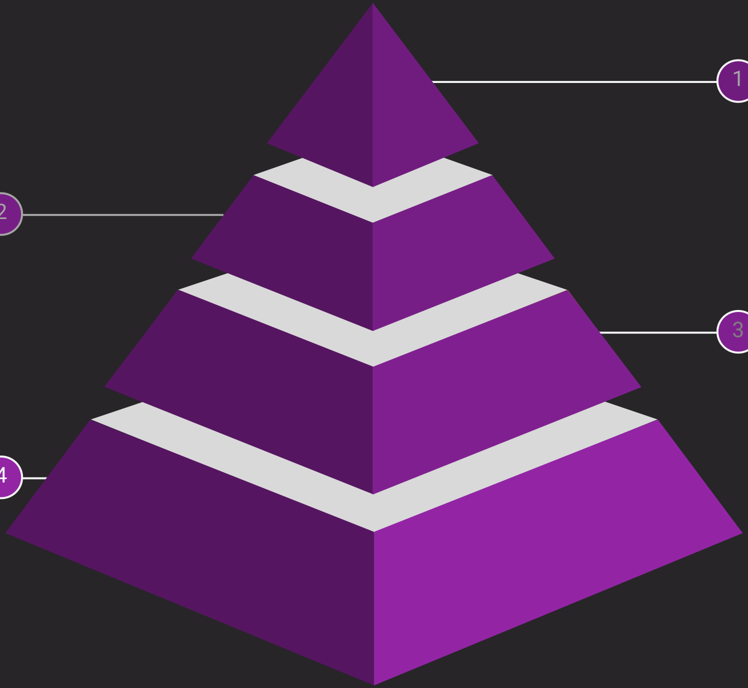
Swish Activation Function
Captures complex patterns and enhances performance.

3. Squeeze and Excitation Optimization

- **Squeeze:** Uses global average pooling to reduce channels to a single number.
- **Excitation:** Highlights the important parts of the channel.



Key Innovations in EfficientNet



1 **Scaling**

Applies Compound Scaling based on a set of fixed coefficients.

2 **MBConv**

Reduces parameters and enhances efficiency through depthwise convolutions.

3 **Squeeze & Excitation**

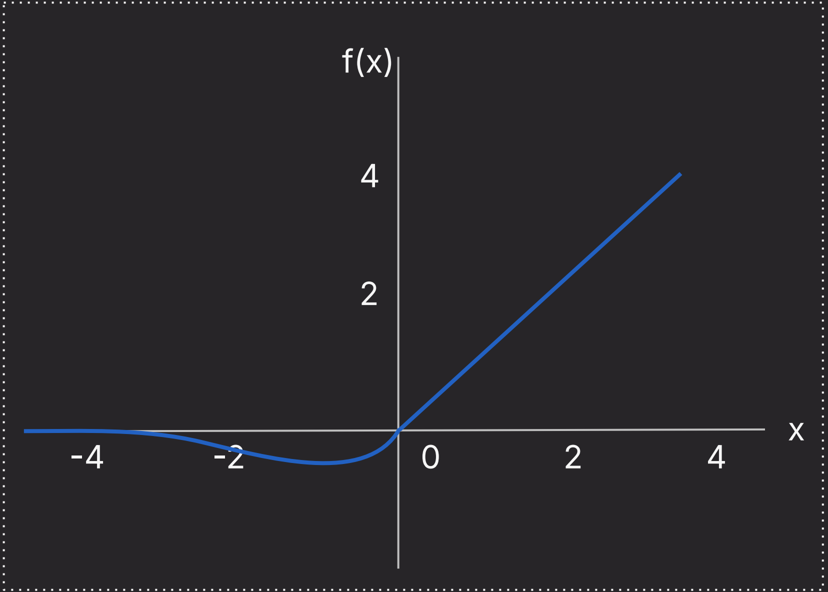
Recalibrates channel-wise features.

4 **Swish Activation Function**

Captures complex patterns and enhances performance.

4. Swish Activation Function

- **Smooth Gradients:** Offers smoother gradients and improved training dynamics.
- **Retains Negative Details:** Allows some negative values to pass through, preserving important details.



Swish (x) = x * sigmoid (x)

EfficientNet

Hands-on

Summary

Model Architecture	Test Accuracy
LeNet	58.41%
AlexNet	53.02%
VGG16 (Feature Extraction Model)	**93.49%**
Inception_v1	88.41%
ResNet50	90.48%
DenseNet121	91.27%
EfficientNet B0	89.21%

Analytics Vidhya

Recent Developments in Computer Vision

2020

ViT (Vision Transformers)

Used Transformer architecture to process sequence of images

2021

MLP-Mixer

Used multi-layer perceptrons to mix features across spatial dimensions

2021

EfficientNet v2

Used a new scaling method for smaller, faster-training models

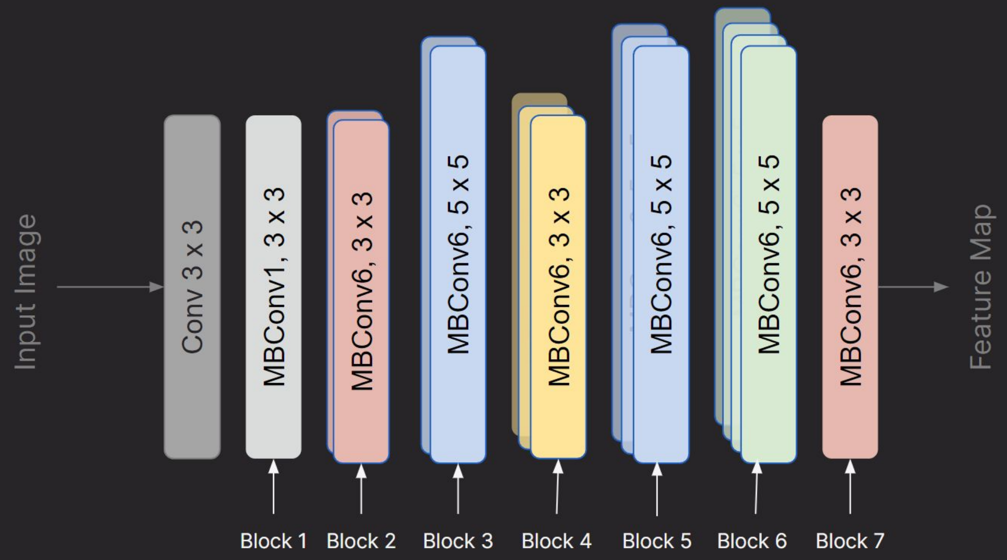
2022

ConvNext

Adapted a transformer-like framework to boost performance in image recognition tasks

EfficientNet Architecture

- Variants B1 to B7 scale uniformly using a compound coefficient.
- Scaling differs from earlier models' random approaches.
- Coefficients derived systematically via AutoML for optimal performance.



EfficientNet Architecture

MBConv Blocks and Squeeze-and-Excitation Blocks:

- Uses MBConv blocks like MobileNetV2, enhanced with squeeze-and-excitation blocks.

Design fine-tunes channel responses, enhancing feature focus in image processing.

MBConv blocks keep EfficientNet lightweight yet enhance performance.

EfficientNet Architecture

Fine tuned Scaling with AutoML MNAS:

- EfficientNet uses AutoML MNAS for performance and resource-aware scaling.
- Variants stay efficient, balancing memory and FLOPs.
- CAMs reveal focus areas, providing insights into model priorities in image analysis.

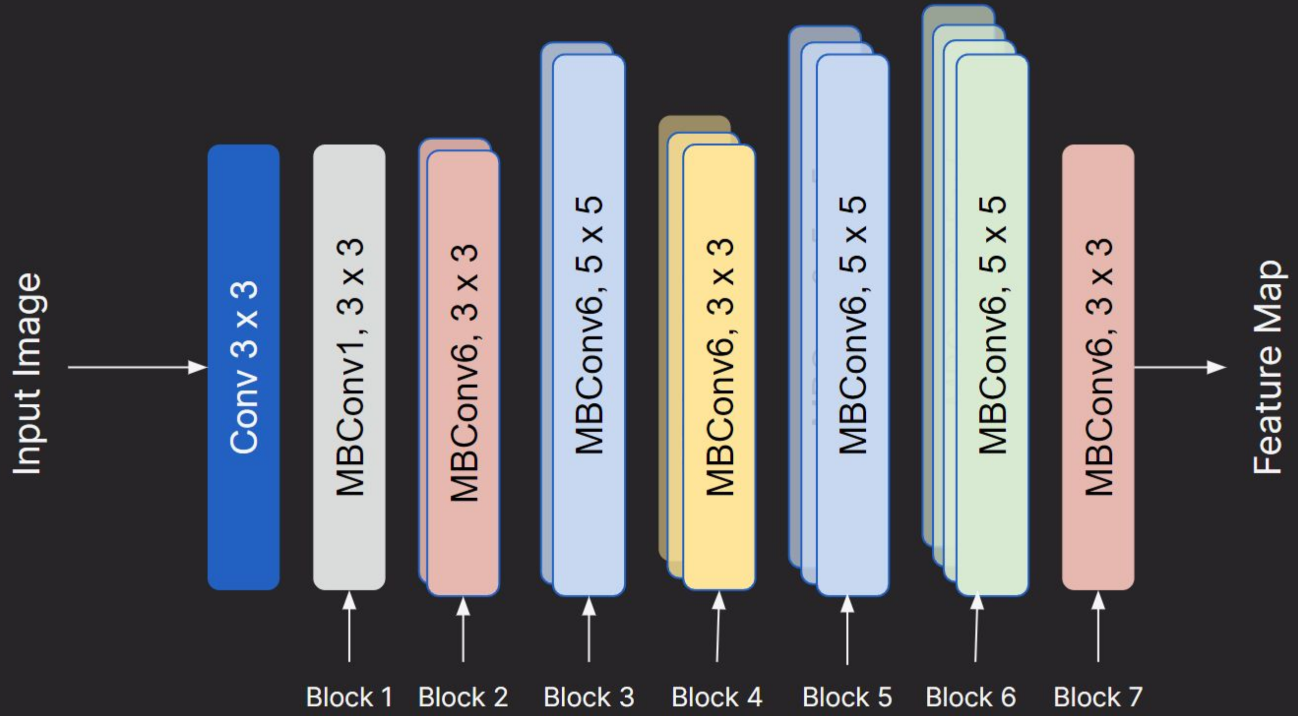
EfficientNet v2

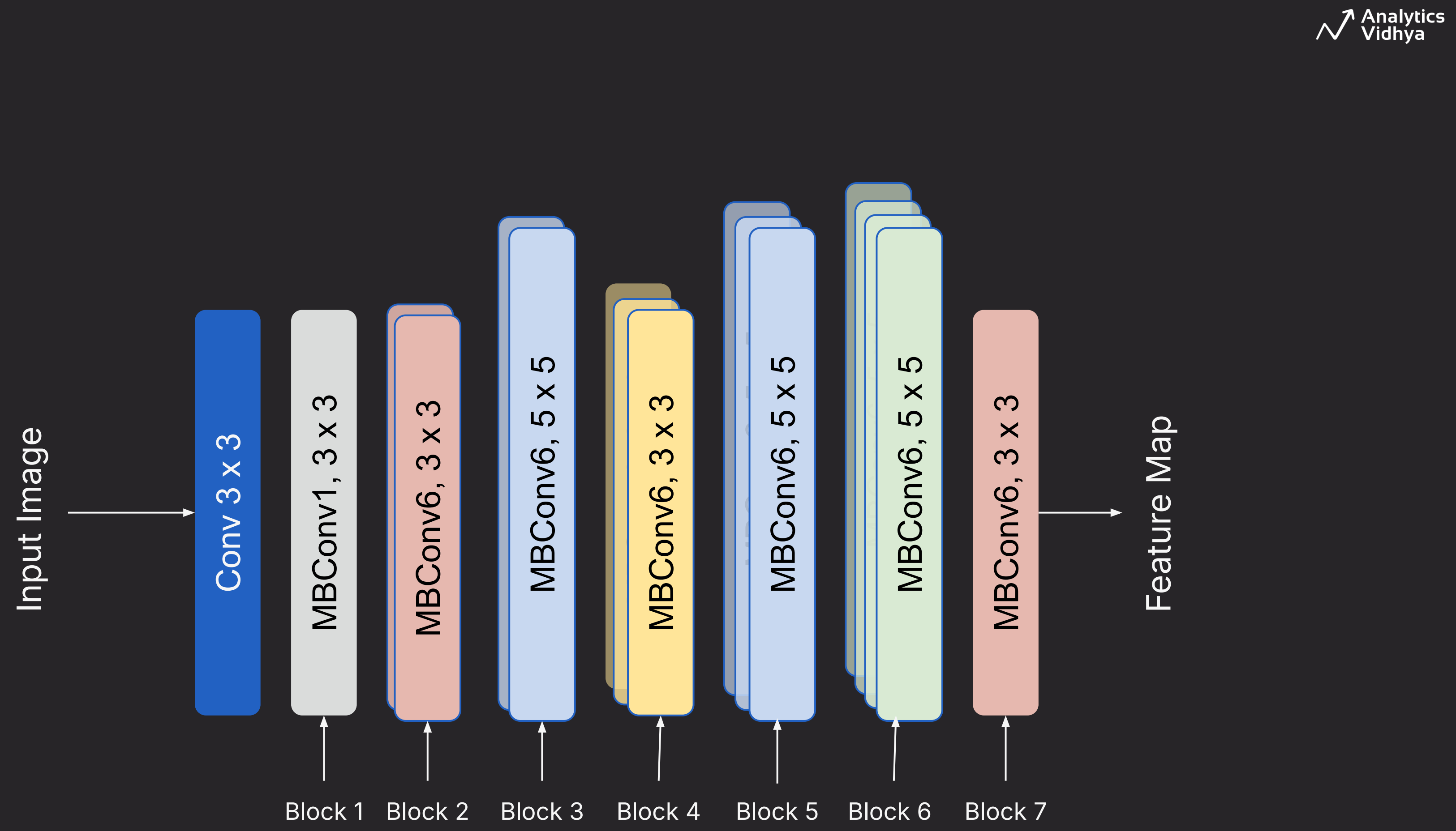
- **Architecture:** Uses a combination of MBConv and Fused-MBConv layers, which are more efficient for both training and inference.

- **Performance:** Achieves better accuracy and efficiency compared to V1, with faster training times and improved robustness to various tasks.

- **Variants:** Includes EfficientNet V2-S (small), V2-M (medium), and V2-L (large), designed to provide a balance between speed and accuracy.

EfficientNet




Input Image
Conv 3 x 3
MBConv1, 3 x 3
MBConv6, 3 x 3
MBConv6, 5 x 5
MBConv6, 3 x 3
MBConv6, 5 x 5
MBConv6, 5 x 5
MBConv6, 3 x 3
Feature Map
Block 1
Block 2
Block 3
Block 4
Block 5
Block 6
Block 7
Analytics Vidhya

EfficientNet

Width Scaling (α) **:** Refers to number of channels in each layer.

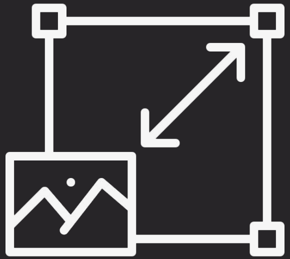
- Model can capture more complex patterns and features.
- Increase in width improves accuracy.

Depth Scaling (β) **:** Refers to total number of layers in network.

- Deeper networks capture more intricate details.
- Demands more computational resources.

EfficientNet

Resolution Scaling (γ) **:** Involves adjusting image input size.



- Higher-resolution images offer more detail, enhancing performance.
- Require increased memory and processing resources.
- Lower-resolution images save resources but lose details.
- Scale resolution by multiplying r by φ^{γ}.

EfficientNet

α β γ

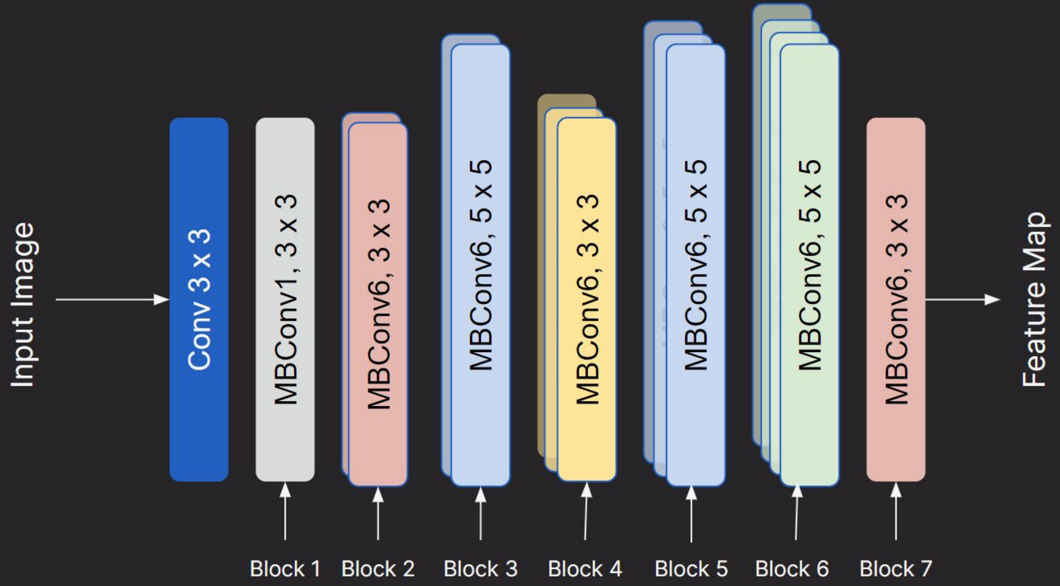
Alpha Beta Gamma

Optimized through AutoML

✓ Best trade-off between Accuracy and Efficiency

EfficientNet

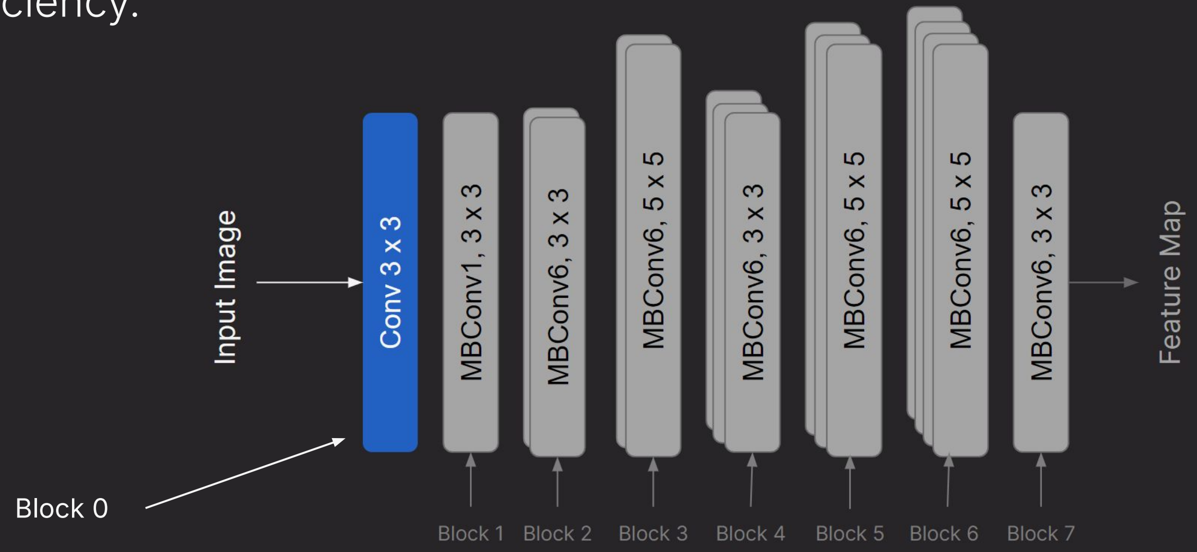
- EfficientNet scales depth, width, and resolution together, optimizing model size, accuracy, and efficiency.
- Balanced scaling reduces computational cost.



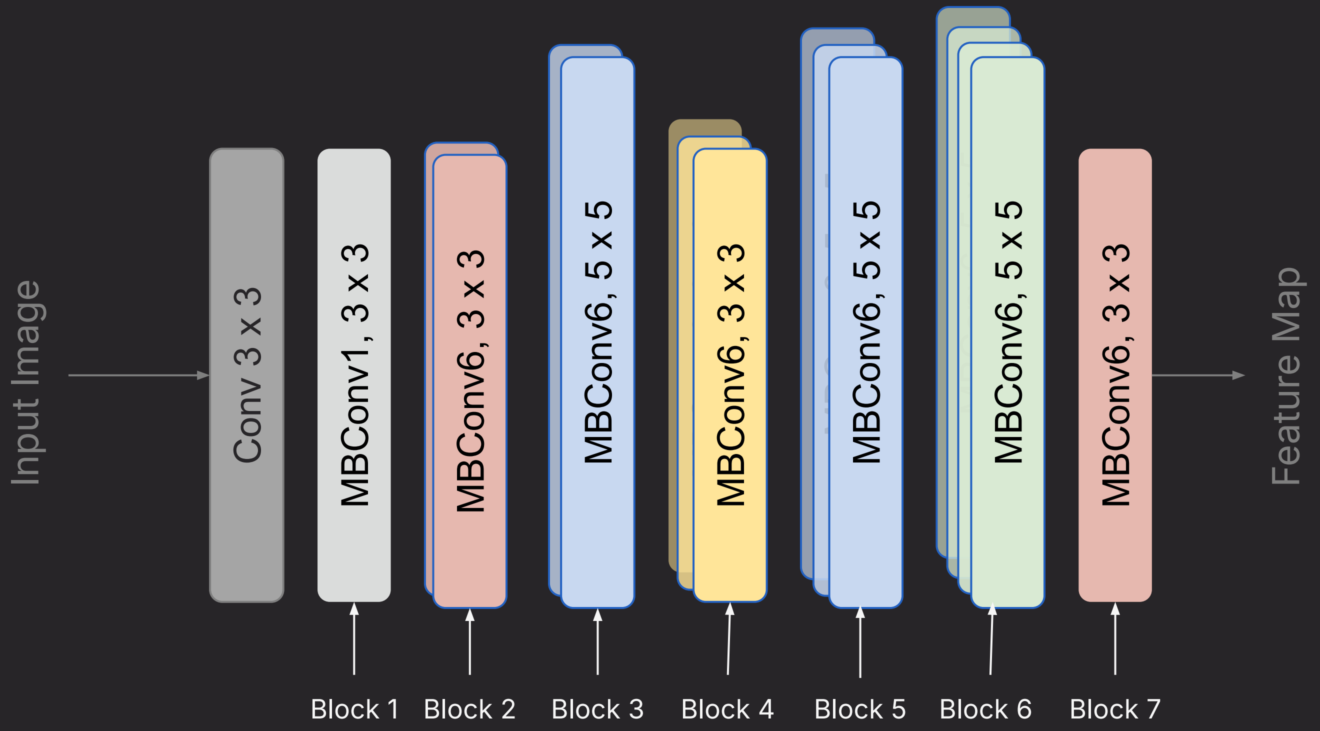
EfficientNet Architecture

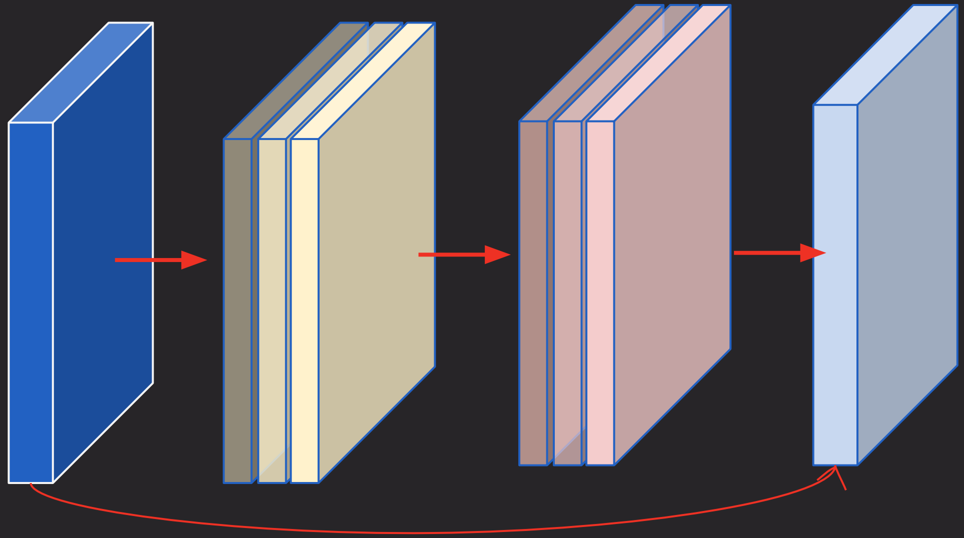
Base Model and Compound Scaling

- EfficientNet-B0: Heart of architecture, designed for optimal accuracy, efficiency.



EfficientNet Architecture





Analytics Vidhya