Deep Learning

Medical Image Classification with ResNets, DenseNets, EfficientNet and ViT

Team: **JAX**

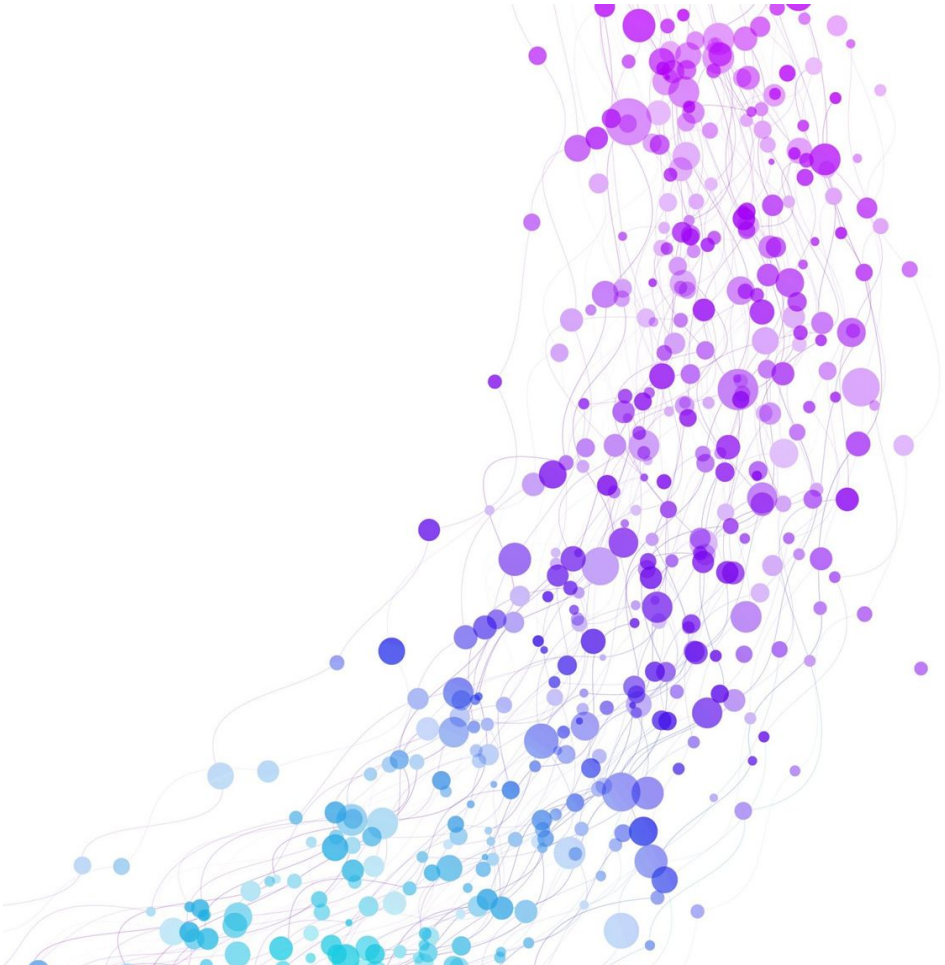
GIT Repositary

Skoltech

Road Map

Skoltech
Skolkovo Institute of Science and Technology

1. Problem statement
2. Datasets description
3. Method
4. Model
    - 4.1. ResNet
    - 4.2. DenseNet
    - 4.3. EfficientNet
    - 4.4. ViT
5. Evaluation Result
6. Conclusion



Problem Statement

Skoltech
Skolkovo Institute of Science and Technology

Medical image classification plays an increasingly important role in healthcare, especially in diagnosing, treatment planning and disease monitoring.

Problem: The lack of large publicly available datasets with annotations means it is still very difficult, if not impossible, to achieve clinically relevant computer-aided detection and diagnosis (CAD).

Deep learning models have been shown to be very effective at image classification, and they are increasingly being used to **improve medical tasks.**

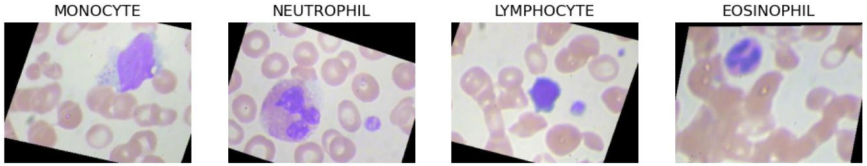
$\Rightarrow$ discuss the implementation of four different deep learning models for medical image classification: ResNet, DenseNet, Efficient Net, and ViT

Datasets Description

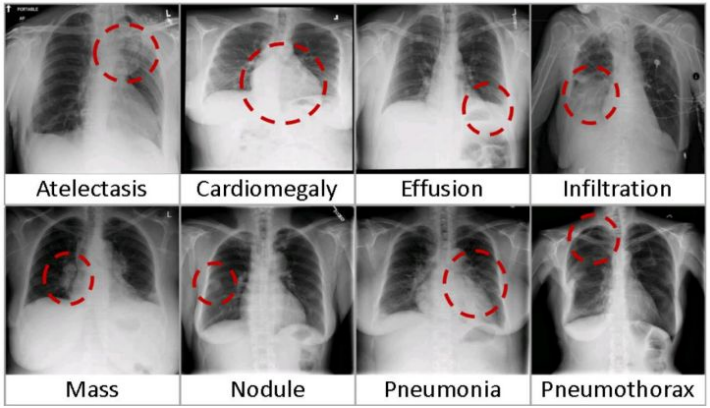
Skoltech
Skolkovo Institute of Science and Technology

Name	Blood Cell Images	Random Sample of NIH Chest X-ray Dataset
Description	12,500 augmented images of blood cells	5,606 chest X-rays from random patients
Number labels	4	15

Blood Cell Images Dataset



NIH Chest X-ray Dataset

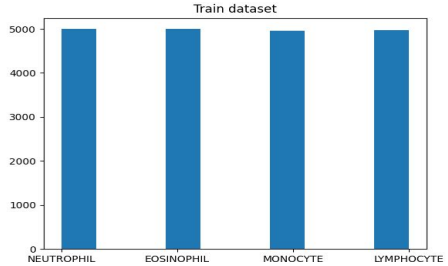


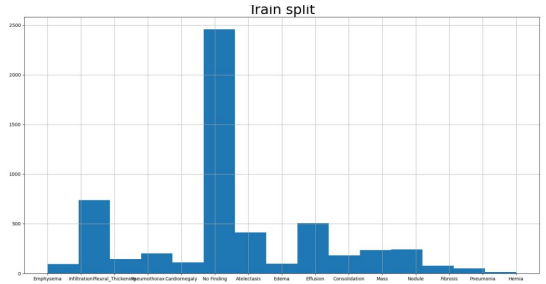
Datasets Description

Skoltech
Skolkovo Institute of Science and Technology

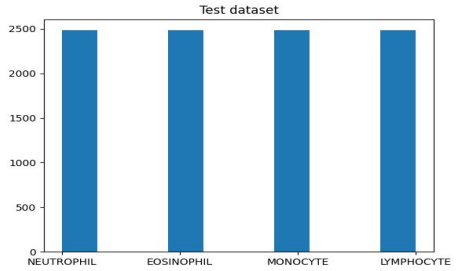
Name	Blood Cell Images	Random Sample of NIH Chest X-ray Dataset
Description	12,500 augmented images of blood cells	5606 chest X-rays from random patients
Number labels	4	15

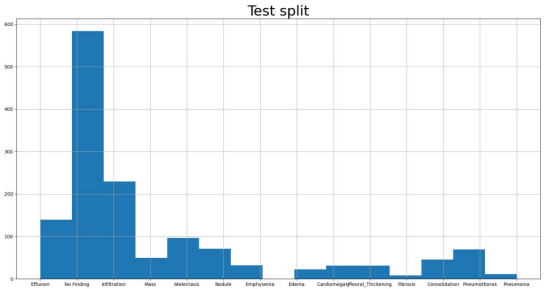
Train split:





Test split:

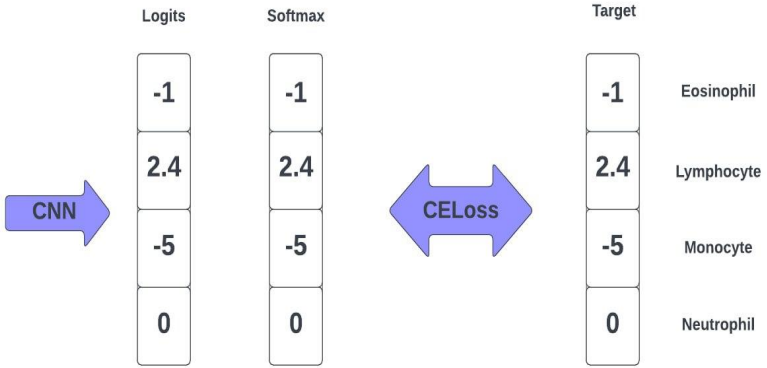




Method

Multi-class classification

Blood Cell Images

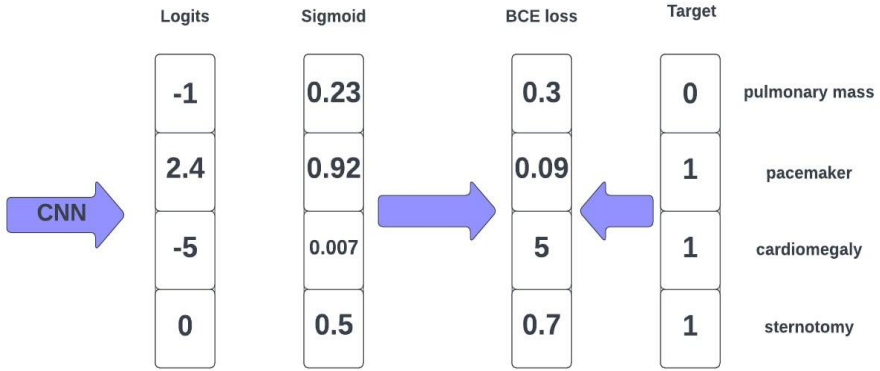


$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \qquad CE = -\sum_i^C t_i log(f(s)_i)$$

each instance belongs to only one class

Multi-label classification

Random Sample of NIH Chest X-ray



$$\mathrm{BCELoss} = -\frac{1}{N}\sum_{i=1}^{N}\left[\mathrm{y}_i \cdot \log\left(p\left(\mathrm{y}_i\right)\right) + \left(1-\mathrm{y}_i\right) \cdot \log\left(1-p\left(\mathrm{y}_i\right)\right)\right]$$

each instance can belong to multiple labels simultaneously

ResNet - Residual Network

ResNets work by stacking together a series of residual blocks.

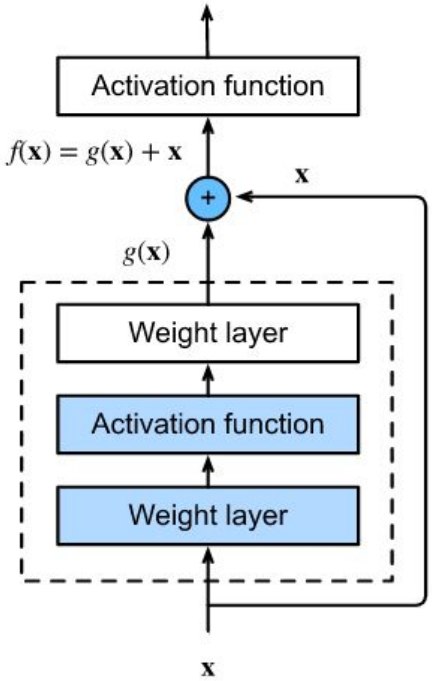
Each **residual block** consists of

- **convolutional layers**
- followed by a **skip connection**

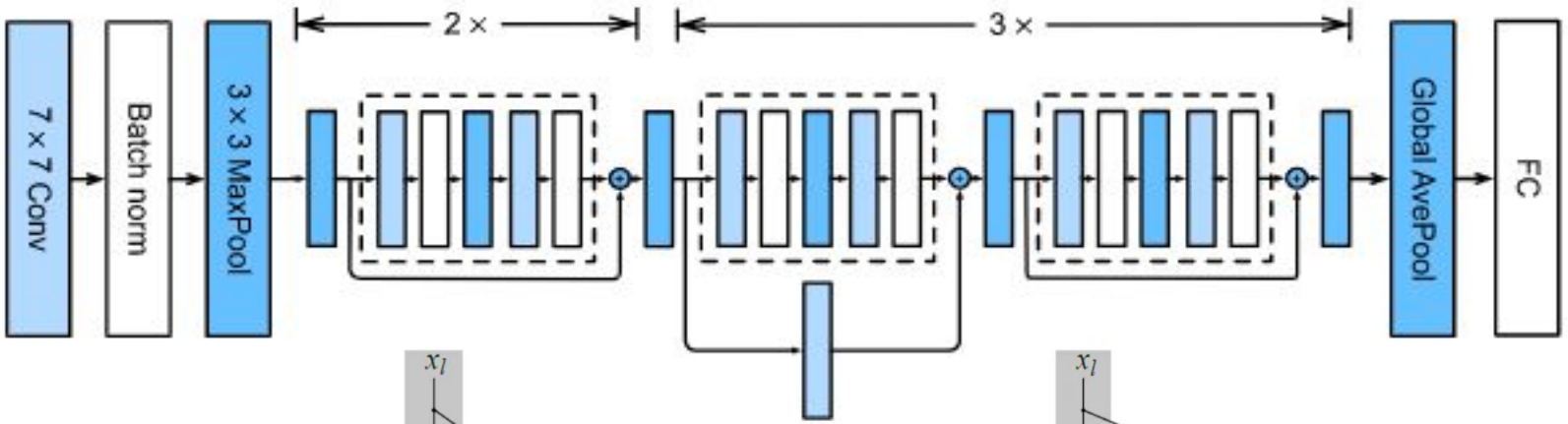
The skip connection allows the output of the first convolutional layer to be added to the output of the final convolutional layer.

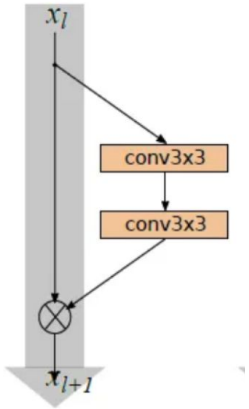
⇒ prevent the *vanishing gradient* problem

of the very deep neural networks

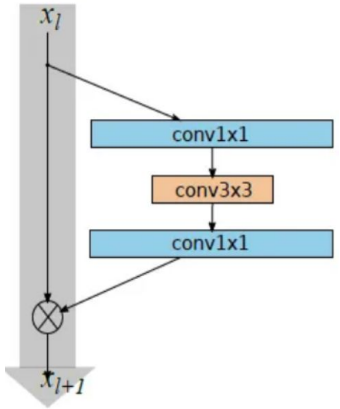


ResNet: Architecture





(a) basic

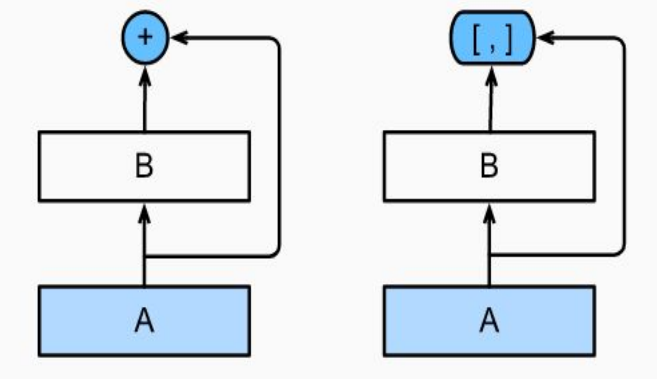


(b) bottleneck

ResNet18, ResNet34

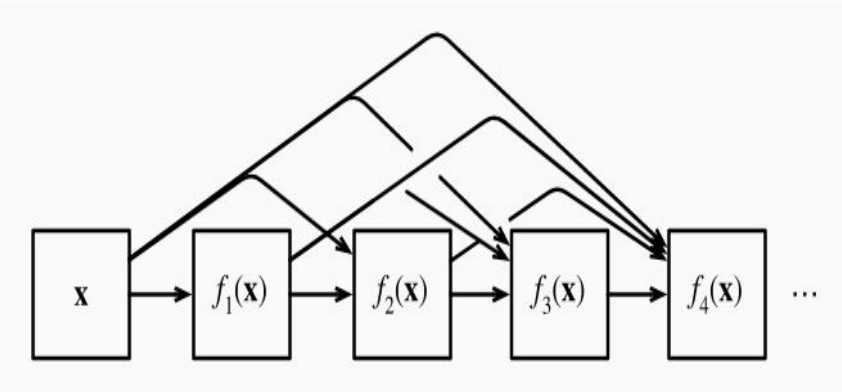
ResNet50, ResNet101, ResNet152

DenseNet - Densely Connected Networks



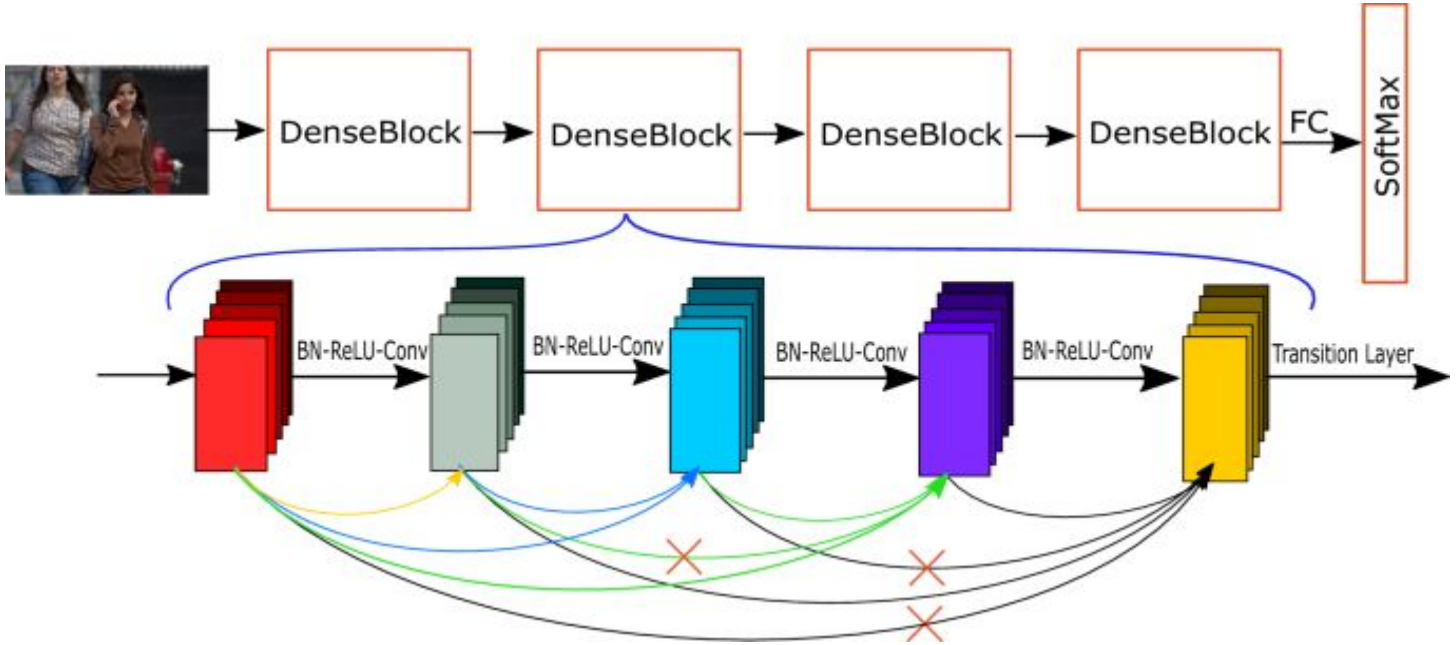
ResNet DenseNet

The main difference between ResNet and DenseNet in **cross-layer connections** is the use of addition vs concatenation



DenseNets are characterized by their use of dense connections

DenseNet - Densely Connected Networks



The main components of a DenseNet are

- dense blocks
- transition layers

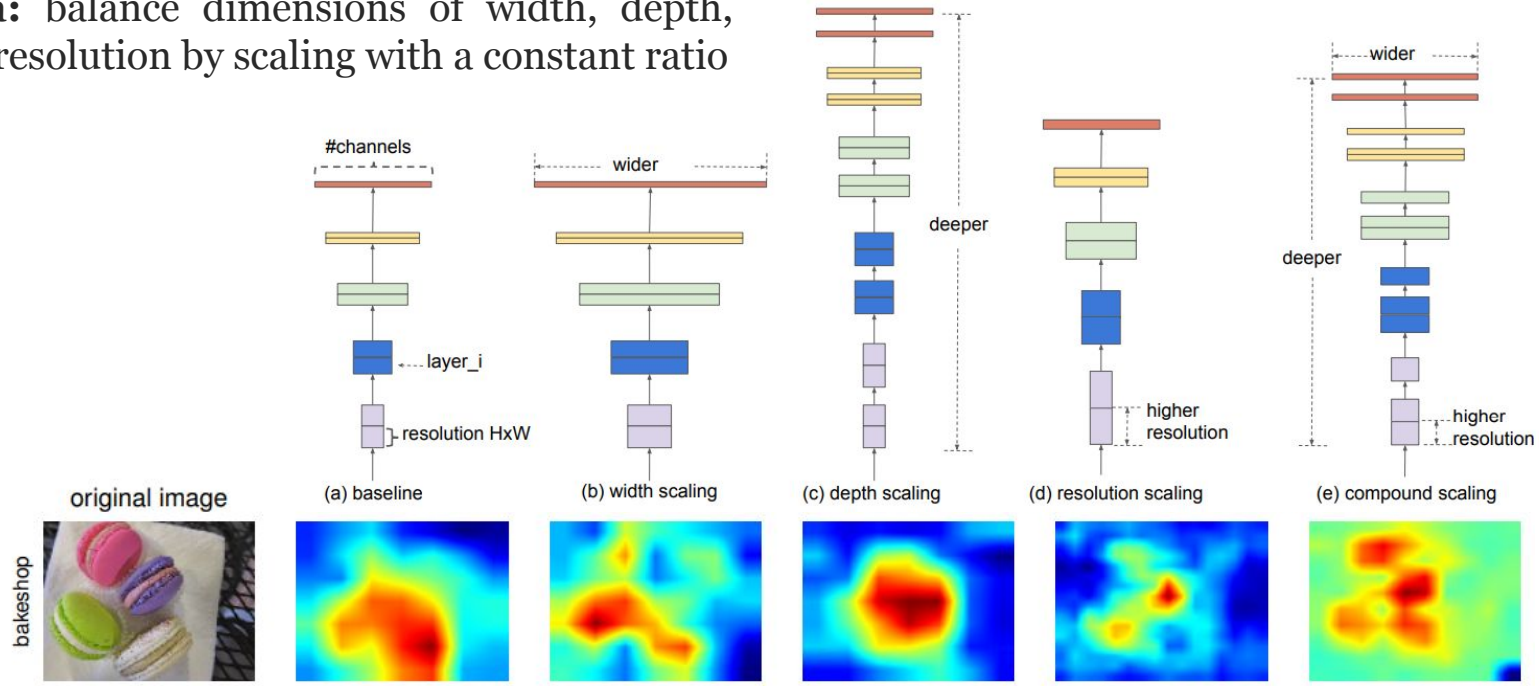
$\Rightarrow$ allow information to flow more easily between different layers of the network

EfficientNet

Skoltech
Skolkovo Institute of Science and Technology

Traditional approaches for a ConvNet scale: width, depth and resolution scaling

Idea: balance dimensions of width, depth, and resolution by scaling with a constant ratio



EfficientNet

Skoltech
Skolkovo Institute of Science and Technology

ConvNet - a set of blocks of layers with input shape as (H, W, C).

$$\max_{d,w,r} \quad Accuracy\big(\mathcal{N}(d,w,r)\big)$$

$$s.t. \quad \mathcal{N}(d,w,r) = \bigodot_{i=1\ldots s} \hat{\mathcal{F}}_i^{d\cdot\hat{L}_i}\big(X_{\langle r\cdot\hat{H}_i, r\cdot\hat{W}_i, w\cdot\hat{C}_i\rangle}\big)$$

$$\text{Memory}(\mathcal{N}) \le \text{target\_memory}$$

$$\text{FLOPS}(\mathcal{N}) \le \text{target\_flops}$$

Compound scaling method

$$\text{depth: } d = \alpha^\phi$$

$$\text{width: } w = \beta^\phi$$

$$\text{resolution: } r = \gamma^\phi$$

$$\text{s.t. } \alpha\cdot\beta^2\cdot\gamma^2 \approx 2$$

$$\alpha \ge 1, \beta \ge 1, \gamma \ge 1$$

Architecture of EfficientNet B0

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Note: use Stochastic Depth with 0,06

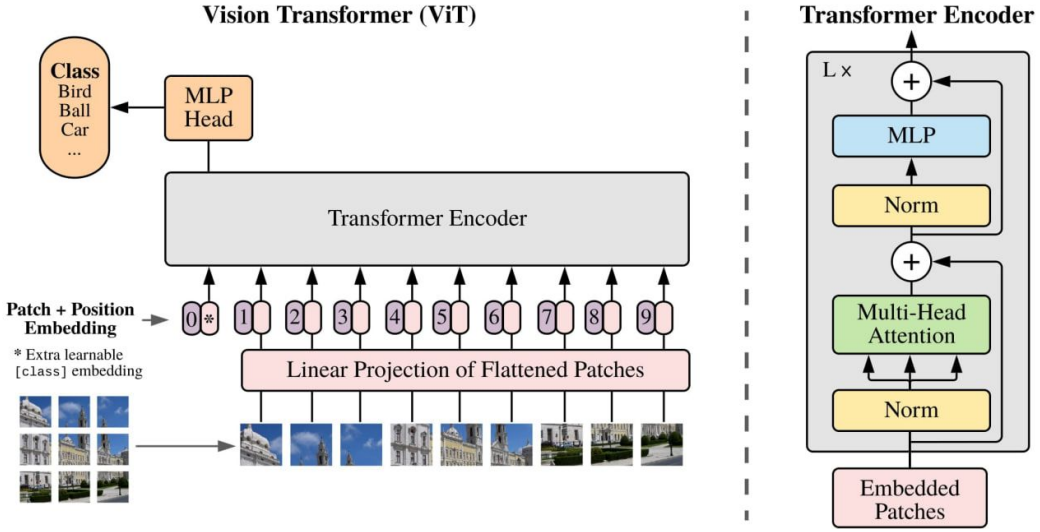
→ EfficientNet family of ConvNet

Vision Transformer (ViT)

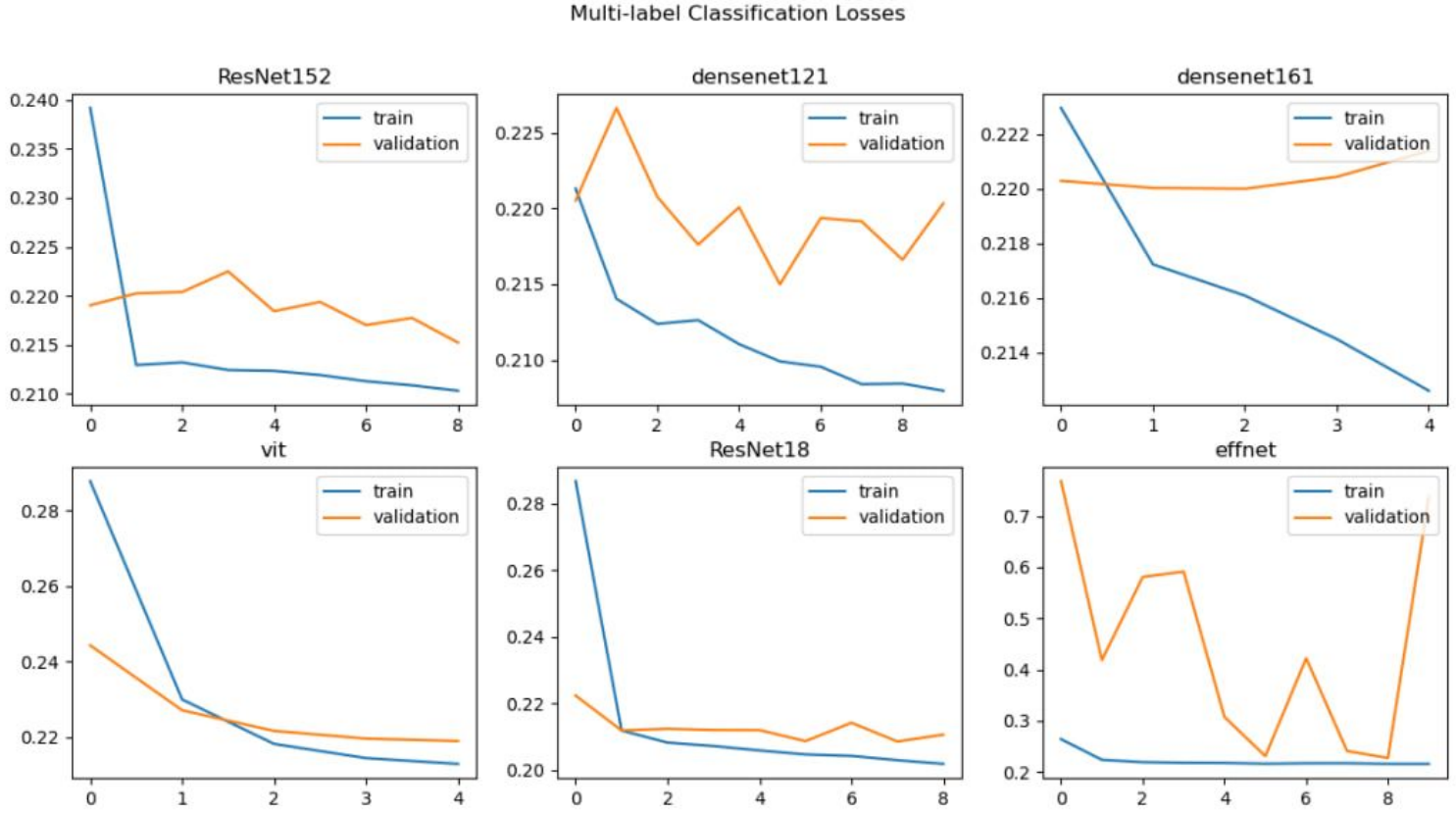
Skoltech
Skolkovo Institute of Science and Technology

Transformer hyperparameters:

Patch size: (8, 8),
Model dim: 128,
Depth: 6,
N heads: 16,
Mlp ratio: 4,
Dropout ratio : 0.3,
Attention dropout ratio: 0.3



Evaluation: Multi-class classification Losses



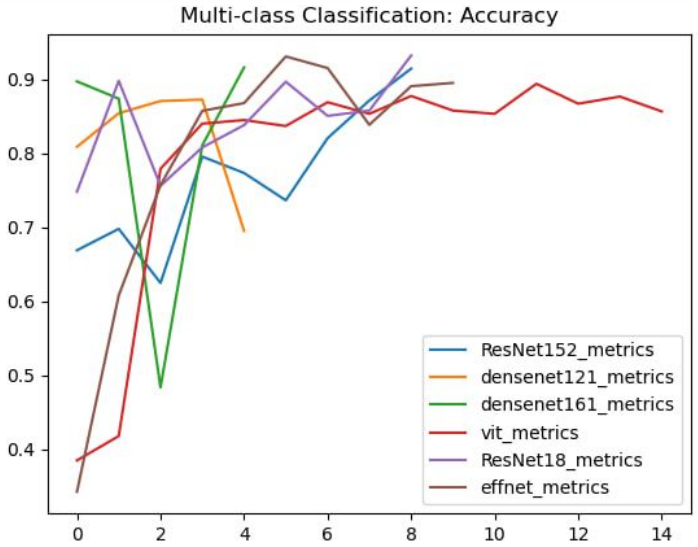
Evaluation: Multi-class classification

Skoltech
Skolkovo Institute of Science and Technology

Multi-class classification metrics

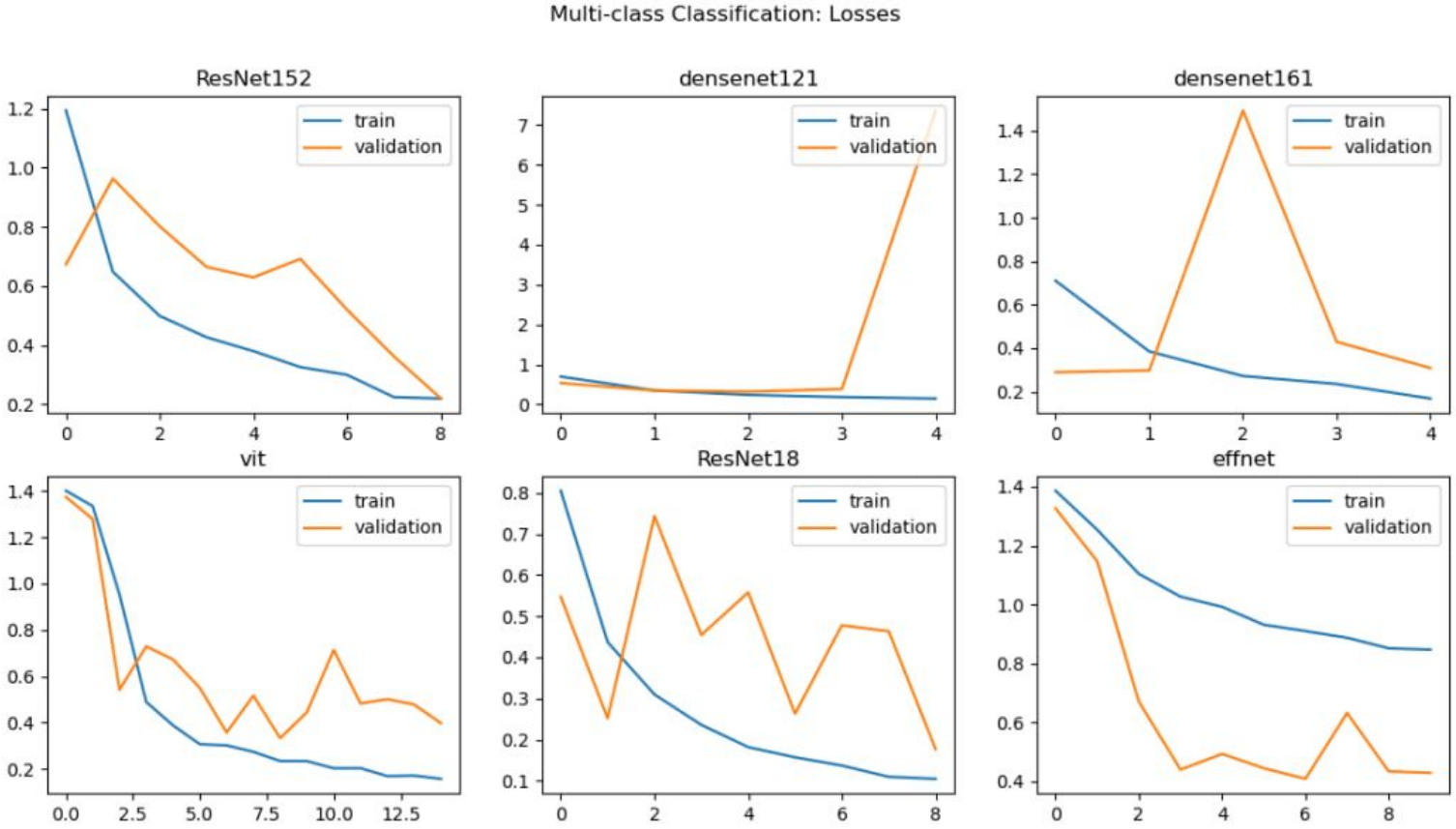
model	accuracy	precision (macro)	precision (micro)	recall (macro)	recall (micro)	F1 (macro)	F1 (micro)
resnet152	0.914757	0.913539	0.914757	0.914980	0.914757	0.913517	0.914757
resnet18	0.932449	0.936440	0.932449	0.932577	0.932449	0.932841	0.932449
densenet121	0.872939	0.887754	0.872939	0.872932	0.872939	0.874157	0.872939
densenet161	0.916365	0.917669	0.916365	0.916586	0.916365	0.915667	0.916365
effnet	0.931242	0.933620	0.931242	0.931398	0.931242	0.931379	0.931242
vit	0.894250	0.912612	0.894250	0.894179	0.894250	0.895673	0.894250

- ResNet18 and Efficient_b0 are the best model with 0.93 accuracy and F1 score
- DenseNet121 and DenseNet161 are unstable with accuracy and F1 are 0.87 and 0.91 for both nets.
- ViT showed SOTA results (accuracy: 0.89, F1: 0.89) with stable learning



Evaluation: Multi-label classification Losses

Skoltech
Skolkovo Institute of Science and Technology



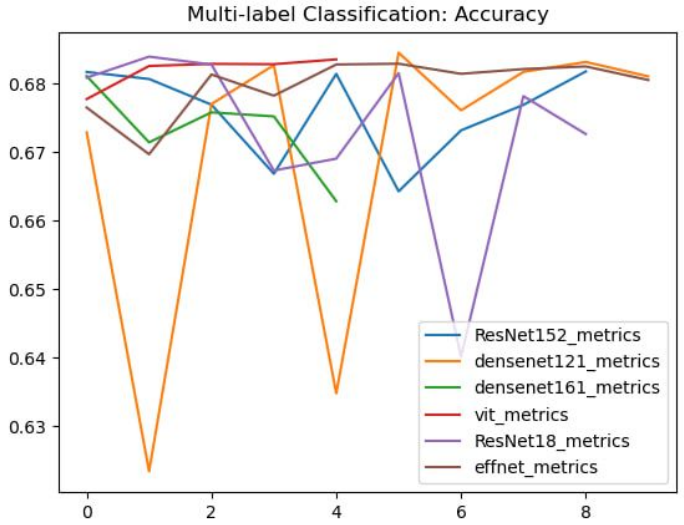
Evaluation: Multi-label classification

Skoltech
Skolkovo Institute of Science and Technology

Multi-label classification metrics

model	Exact Match Ratio	Hamming loss	Recall	Precision	F1
ResNet152	0.501338	0.501338	0.502230	0.501933	0.502052
ResNet18	0.473684	0.473684	0.476360	0.474874	0.475320
densenet121	0.513827	0.513827	0.513827	0.513827	0.513827
densenet161	0.520963	0.520963	0.520963	0.520963	0.520963
effnet	0.520963	0.520963	0.520963	0.520963	0.520963
vit	0.520963	0.520963	0.520963	0.520963	0.520963

- DenseNets and Efficient_b0 showed satisfied results with 0.52 accuracy and F1 score
- DenseNets and ResNets were unstable, while EfficientNet_b0 showed stable learning
- ViT with metrics, as other algorithms and showed stable learning.



Conclusion

Skoltech
Skolkovo Institute of Science and Technology

Number of parameters in models

model	total parameters, ths.	trainable parameters, ths.
resnet_18	11178	11178
resnet_152	58152	58152
densenet_121	6957	6957
densenet_161	26480	26480
effnet	5288	5288
transformer	1368	1368

EfficientNet is the most efficient model for both multi-class and multi-label classification problems. It performs well in all metrics and shows stable learning. ViT performs as other algorithms in terms of metrics. However, there are significantly less params to train. That makes this architecture to be alternative for ConvNets in the future.

Work distribution

Skoltech
Skolkovo Institute of Science and Technology

Team: JAX

Xavier Aramayo Carrasco	Julia Kudryavtseva	Vo Ngoc Bich Uyen
33 %	33 %	33 %
• Find Dataset for Multi-class classification • Implementation of DenseNet	• Implementation of EfficientNet, ViT • Evaluation	• Find Dataset for Multi-label classification • Implementation of ResNet

Thank You!

Skoltech