Kandinsky Patterns - Challenge 1

W. Kosiuk, M. Mazuryk, B. Sawicki

Wydział MiNI
Politechnika Warszawska

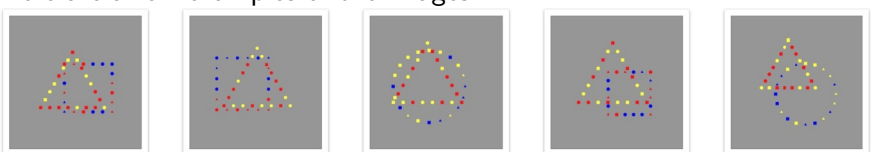
May 25, 2022

Presentation plan

1. Challenge description
2. Interpretability for image classification
3. Methods and related work
4. Trained models
5. Results
6. Comparison of explanations
7. Conclusions

Challenge Description

We chose Challenge 1 as the one we are working on. Dataset included in this challenge contains images with small objects that are arranged on big shapes in such way that big shapes are always different than the small ones.
Here are a few examples of the images:



- 3 shapes - circle, square,triangle
- 3 colors - blue, yellow, red
- 3 types of images - ground truth, not belonging images, false images

Interpretability

Interpretability of models is a really important part of creating a machine learning model. In our project we decided to focus on 2 areas of explanations:

Shape

Explanations about the shapes

Random class explanation

Image targets are randomized so model should not learn

Methods and Related Work

Algorithms used for explanations

- LIME
- SHAP
- Saliency
- Integrated Gradients
- InputXGradient
- GradientSHAP

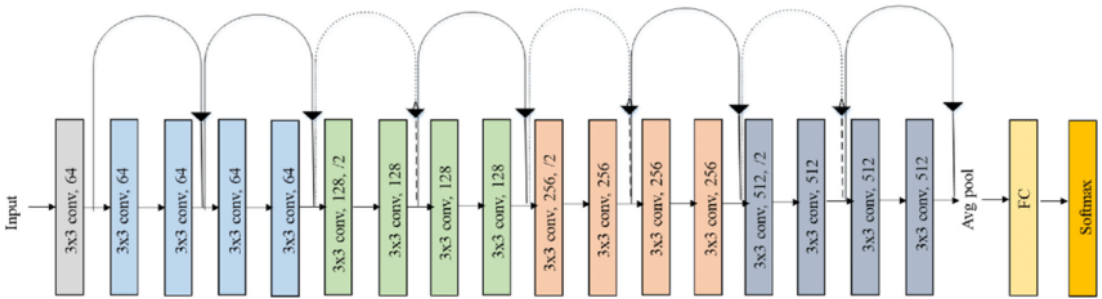
Related work:

- H. Müller A. Holzinger - Kandinsky Patterns
- J.Adebayo, J.Gilmer, M.Muelly, I.Goodfellow, M.Hardt, B.Kim - Sanity Checks for Saliency Maps

ResNet18

ResNet18

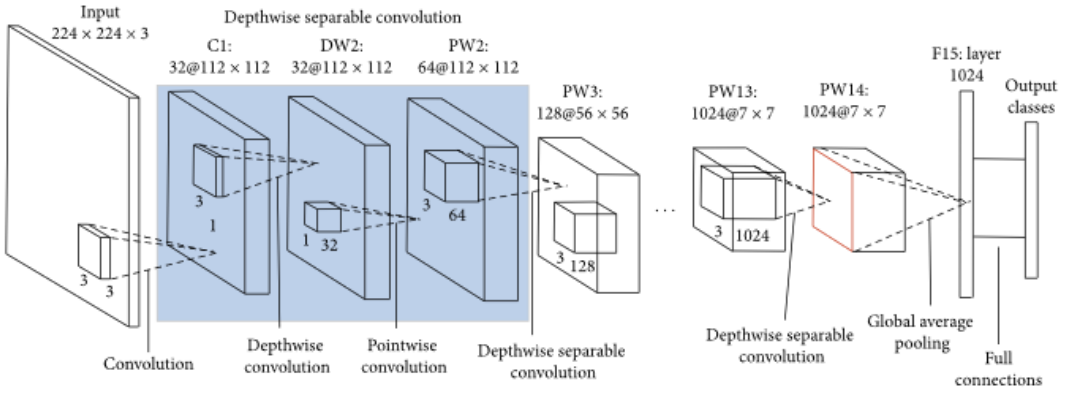
18 layers deep Convolutional Neural Network



MobileNet

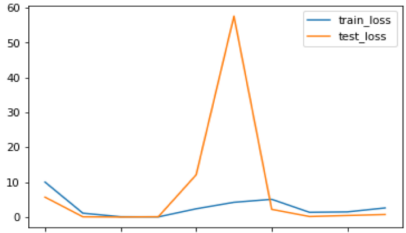
MobileNet

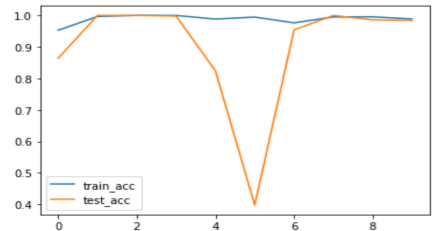
Small, low-latency, low-power model with reduced number of parameters



Results

ResNet18 results:

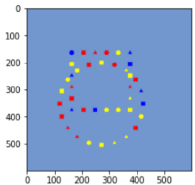


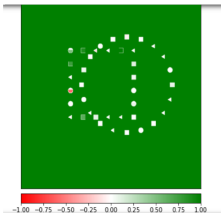


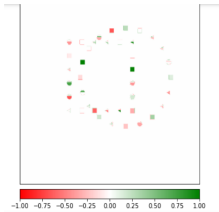
MobileNet results:
Pre-trained MobileNet
with 10 epochs of training
Accuracy: 0.986

Comparison of Explanations

MobileNet:







Conclusions

The End