Artistic Insights: Classifying and Interpreting Artwork Styles Using Deep Learning

Course: Applied Deep Learning
Student: Viktoriia Ovsianik [12217985]

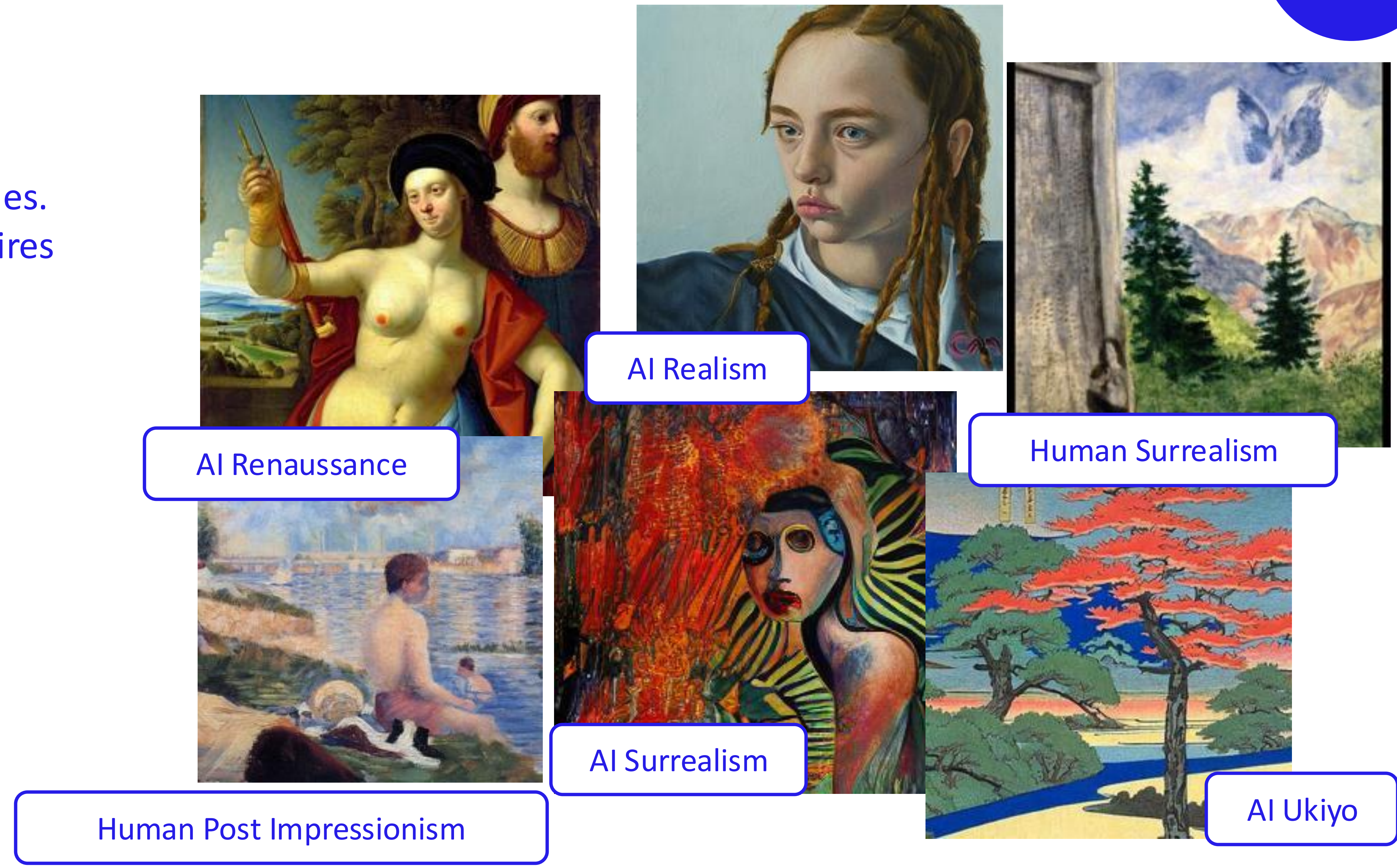
What is this project about?

- **Goal :**
  Classify artwork images to identify styles and distinguish if they were generated by AI or created by humans.

- **Why It Matters :**

  - Difficulty in differentiating overlapping artistic styles.
  - Increasing presence of AI-generated artwork requires robust classification tools.

Will you be able to say if this artwork was generated by AI or created by human? And what about the style?



Dataset Overview

STRUCTURE:

180.000+

Images of paintings

60.000

Created by humans

[14th to the 21st century]

The rest – AI Generated:

- Latent Diffusion
- Standard Diffusion

Artworks of the following styles:

- Art nouveau
- Baroque
- Expressionism
- Impressionism
- Ukiyo-e
- Post impressionism
- Realism
- Renaissance
- Romanticism
- Surrealism

Dataset was extracted from Kaggle: Link to the dataset

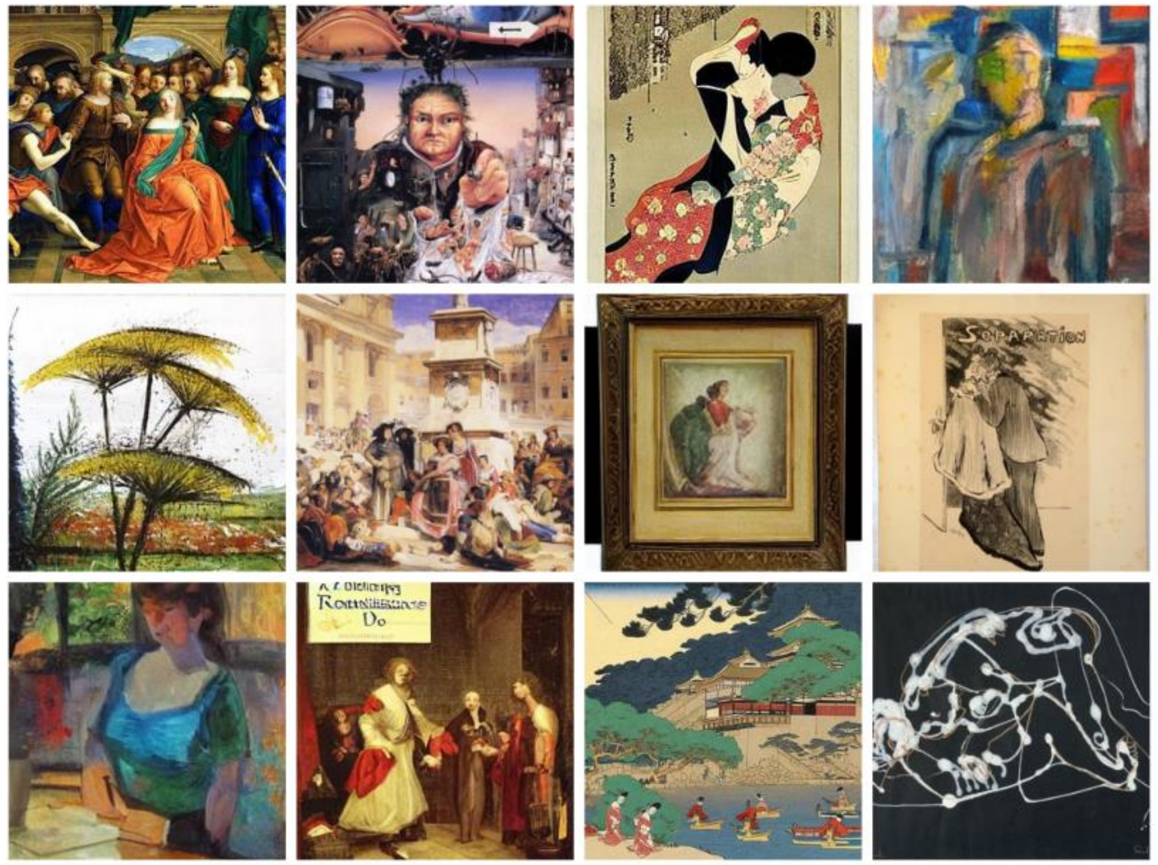
PREPROCESSING

1. Balancing classes in train set (5.000 each)
2. Resize images to 256x256
3. Train/validation split (90% : 10%)

90k/10k/30k

(train/validation/test)

4. Labels: {AI/human}_{artistic_style_name}



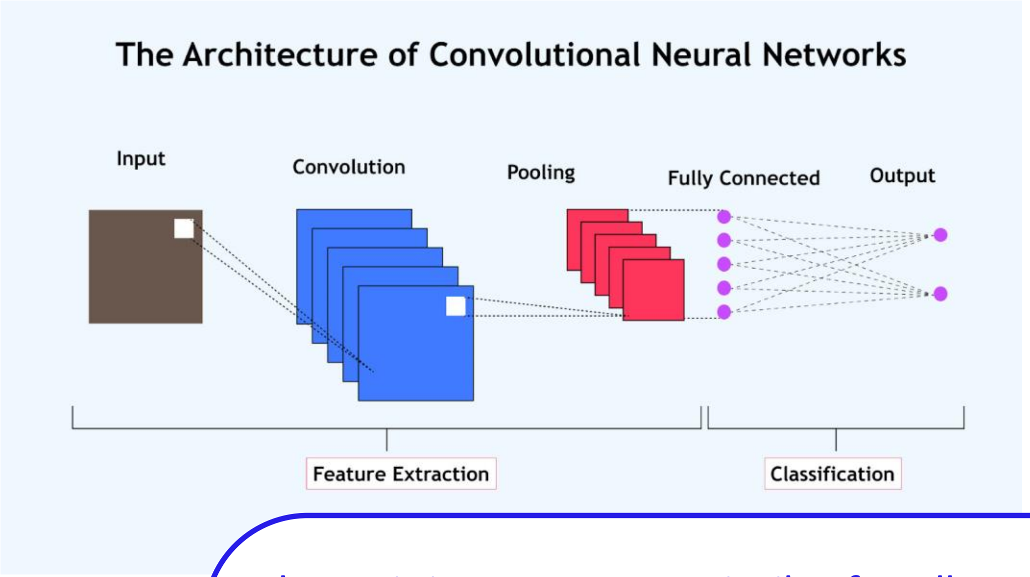
Models

Simple CNN

A **lightweight convolutional neural network** designed for image classification.

It features:
- Two convolutional layers with ReLU activation
- Max pooling
- A fully connected layer with 20 output units

Additionally, images resized to 32 x 32



Source

The training setup was similar for all models:

- **epochs:** 10
- **patience:** 3
- **validation frequency:** 1
- **learning rate:** 0.001
- **optimizer:** SGD Optimizer
- **loss function:** Cross entropy loss function
- **training resources:** Google Colab GPU

ResNet18

A **residual neural network** designed for image classification. It features:

- A 7x7 convolutional layer with ReLU activation, followed by max pooling.
- Four groups of residual blocks
- A global average pooling layer to reduce spatial dimensions.
- A fully connected layer mapping to 20 output units

Advantages:
- Residual blocks with skip connections allow gradients to flow back to earlier layers.
- Can learn more abstract and complex features.
- Available weights pretrained on larger datasets

ResNet18: base

- Fine-tuned ResNet18 model (from torchvision.models)
- Layers ['conv1', 'bn1', 'layer1', 'layer2'] were frozen during training to focus on deeper feature extraction.
- Images resized to 32 x 32

! Overfitting around 4th epoch

ResNet18: with augmentation

- + Data augmentation: Random Horizontal Flip, Random Rotation(30), Random Resized Crop (32), ColorJitter

! Even worse results

- All layers unfrozen
- Data augmentation: Random Horizontal Flip, Random Vertical Flip, Random Rotation(30)

! Comparable results with ResNet18 base

ResNet18: best performance

- All layers unfrozen
- Data augmentation: Random Horizontal Flip, Random Vertical Flip, Random Rotation(30)
- Images resized to 224 x 224

! Significantly better performance

Results

For test set

1 Simple CNN: baseline

Metric	Value
Test Accuracy	45.12%
Overall Precision	44.83%
Overall Recall	45.12%
Overall F1-Score	44.03%

Accuracy:
+ 18,24 %

2 ResNet18: base

Metric	Value
Test Accuracy	63.36%
Overall Precision	63.60%
Overall Recall	63.36%
Overall F1-Score	63.21%

Accuracy:
+ 3,68 %

ResNet18: with "aggressive" data augmentation

Lead to worse results

3 ResNet18: with data augme

Metric	Value
Test Accuracy	67.04%
Overall Precision	66.55%
Overall Recall	67.04%
Overall F1-Score	66.56%

Accuracy:
+ 15,86 %

4 ResNet18: best performance

Metric	Value
Test Accuracy	82.90%
Overall Precision	82.87%
Overall Recall	82.90%
Overall F1-Score	82.90%

Further improvement?

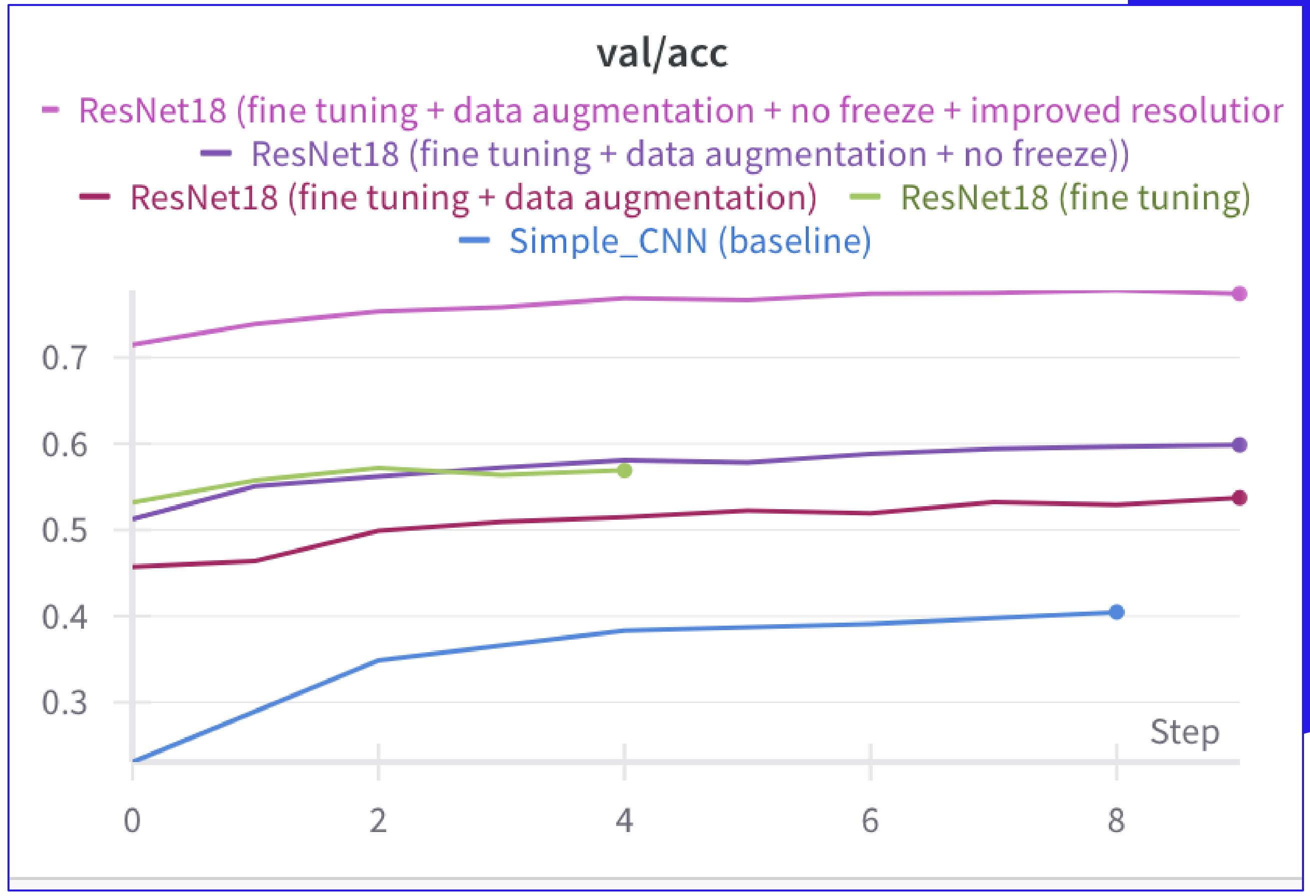


Fig. Accuracy of all trained models.

! Model line colors correspond to model titles on the slide

Key insights & takeaways

- **ResNet50 vs Simpler Models:**
  Many studies focus on ResNet50 but simpler models can also achieve good

- **Impact of Image Resolution on Training:**
  Increasing the image resolution did not significantly extend the training time, yet it resulted in a notable improvement in performance.

- **Data Augmentation and Task-Specific Effects:**
  Not all data augmentation techniques are universally effective.

- **AI-Generated Images Identification:**
  AI-generated images were not hard to identify, would be interesting to consider dataset with more human-like AI generated artworks

App Demo

Let’s now watch a demo of the app :)