

Deep Learning on Kubeflow

Iurii Mozzhorin, Rebekka Pech

July 13, 2019

Hands-on Lab - Big Data Technologies SS 2019

Goethe University Frankfurt am Main

What is Kubeflow?

- an open-source platform
- based on Kubernetes
- simple, portable and scalable deployments of ML
- integration of existing tools and libraries

Includes components for

- model developing, training, serving
- hyperparameter tuning
- pipelines



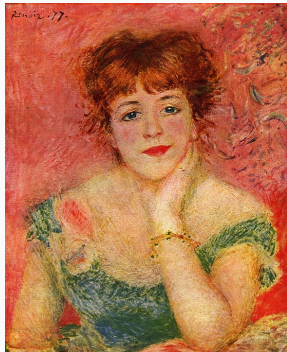
Use case

Goal

- Train a CNN model to distinguish between works of 10 painters

Dataset

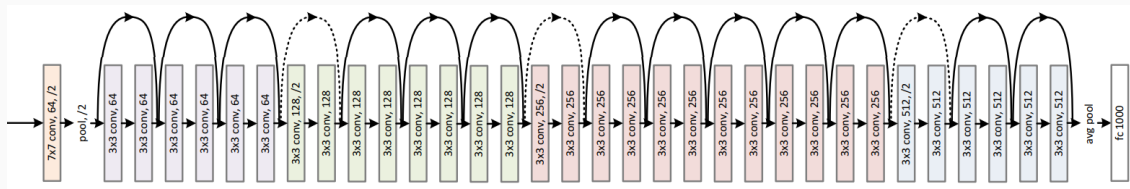
- "Painter by Numbers" from Kaggle.com
- Original size 83GB
- Images dimensions 477 – 7444 px
- Selected & resized:
 - archive 340 MB
 - 388 images \times 10 artists
 - images smallest dimension 448 px
- Test set 20 images (5%)



Use case

Model

- ResNet 50
- Transfer learning with weights from ImageNet (1000 classes)
- Input size 224×224 px
- Image augmentation: shift, rotation, flip



- Preprocess the data
- Create the model (transfer learning)
- Create the Docker image
- Train the model on KF using TFJobs on CPUs
- Save the trained model
- Train the model on KF using TFJobs on GPUs
- Distributed training on multiple pods
- Serve the model

Thank you for your attention!