Università degli Studi di Padova

DIPARTIMENTO MATEMATICA

Transfer learning for crowd numerosity estimation

Pavel Ianko

pavel.ianko@studenti.unipd.it

University of Padua, Data Science MSc

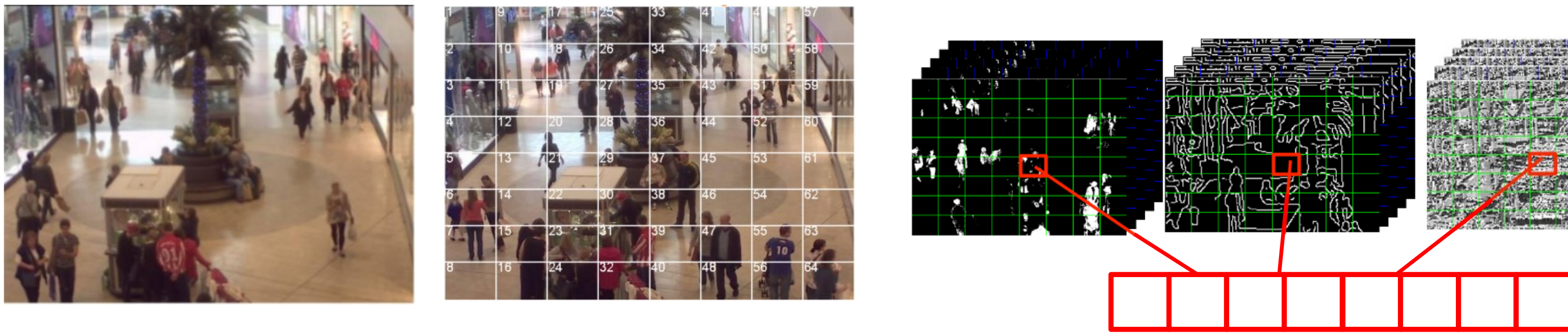
Overview

- Alternative approaches to crowd counting
  - Why ML approach is valid?
- Dataset
  - Target distribution
- Method
  - Selected architectures
  - Target metrics
- Experiments
  - Baseline selection
    (validation MAE, trainable parameters, training time)
  - Pretraining on another dataset
  - Correct pretraining technique
- Test report
- Conclusion

Alternative approaches to crowd counting

Regression-based approach

Idea: learn a regression model on low-level and local visual features



Perspective normalisation map

Cell-splitting

Cell-wise local feature extraction

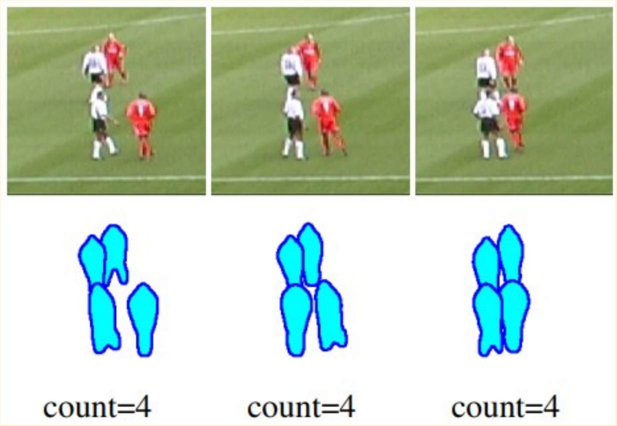
Ordered feature vectors

- Regression model estimates people foreach region independently [1]
- Some approaches learn global regression model [2]
- Created the “Mall Dataset”
- 3.15 MAE

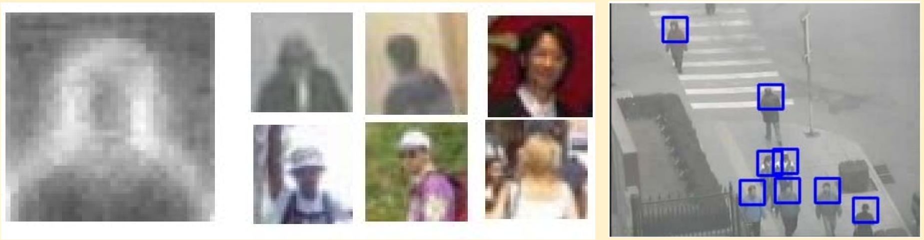
[1]. Ke Chen et al. Feature mining for localised crowd counting.
[2]. A.B. Chan et al. Counting people with low-level features and Bayesian regression.

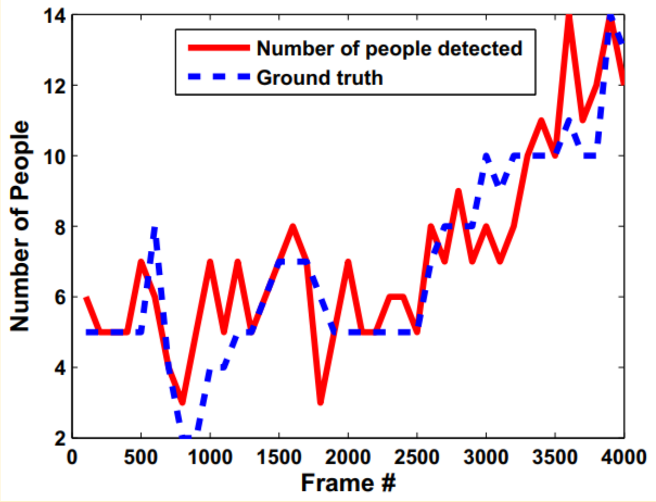
Detection-based approach

Idea: detect instances of people



- Bayesian Marked Point Process model [1].
- Detection rate – 92%



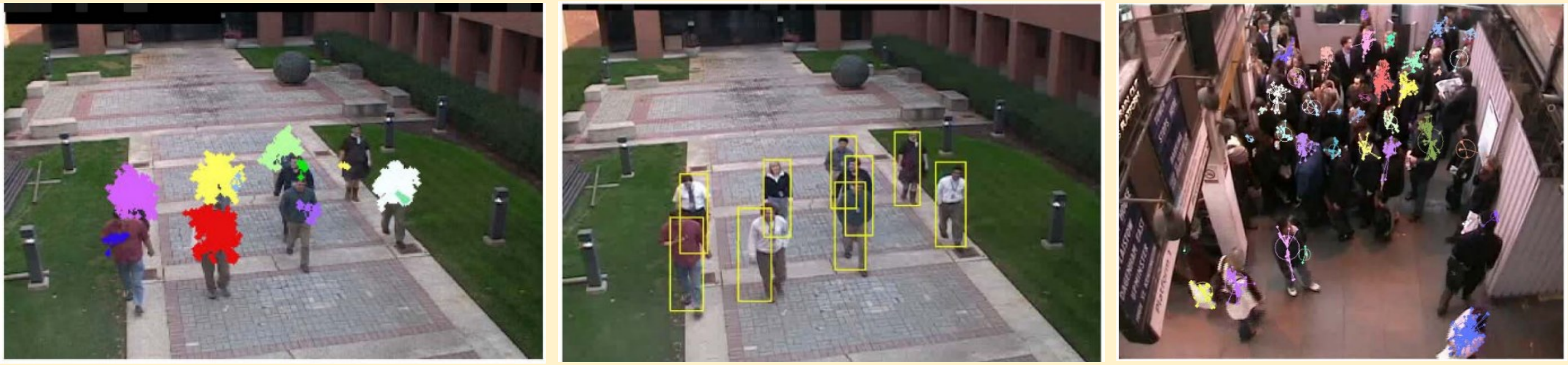


- HOG feature extraction for head-shoulder pattern [2]
- Ada-boost detector [2]

[1]. W. Ge et al. Marked point processes for crowd counting.
[2]. Min Li et al. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection.

Clustering-based approach

Idea: "*pair of points that appears to move together is likely to be part of the same individual*" [1]



- Bayesian clustering model [1]
- **Feature detector** (Rosten-Drummond + Tomasi-Kanade features)
- **Feature tracker**

[1]. G. J. Brostow et al. Unsupervised Bayesian Detection of Independent Motion in Crowds.

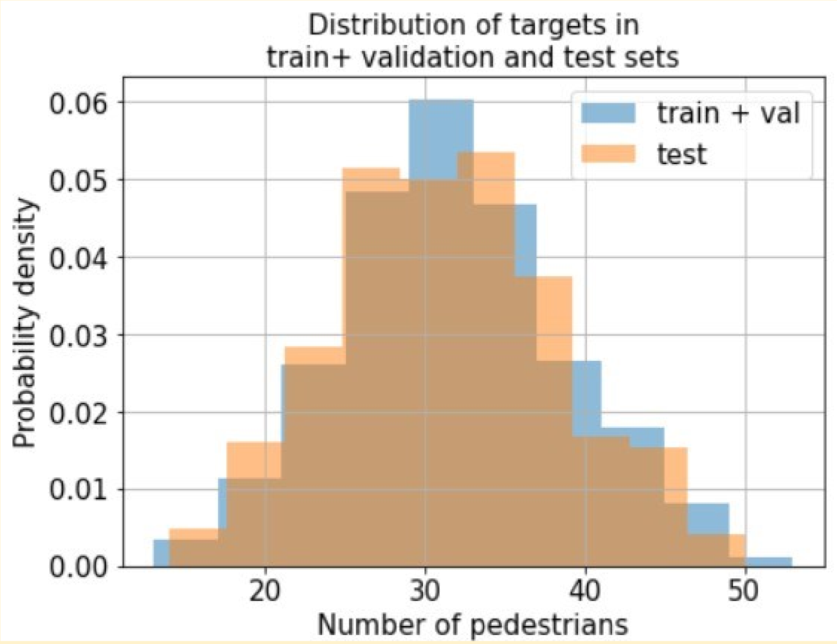
Why CNN approach is valid

	Regression	Clustering	Detection	CNNs
Feature extraction	Manual	Manual	Manual	**Automatic**
Pipeline complexity	>1 components [2, 3, 4, 5]	>1 components [2, 3, 4, 5]	>1 components [2, 3, 4, 5]	**1 model** (end-to-end-learning [1])
Computational efficiency	Yes, compared to clustering and detection [2]	Worse than regression [2]	Worse than regression [2]	Depends on requirements & architecture
Clutter & object occlusion	Performs better [2]	Worse than regression [2]	Worse than regression [2]	Can learn occlusion-robust features

[1]. Andrew Ng. Machine Learning Yearning.
[2]. Ke Chen et al. Feature mining for localised crowd counting.
[3]. G. J. Brostow et al. Unsupervised Bayesian Detection of Independent Motion in Crowds.
[4]. Min Li et al. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection.
[5]. A.B. Chan et al. Counting people with low-level features and Bayesian regression.

Dataset

Dataset



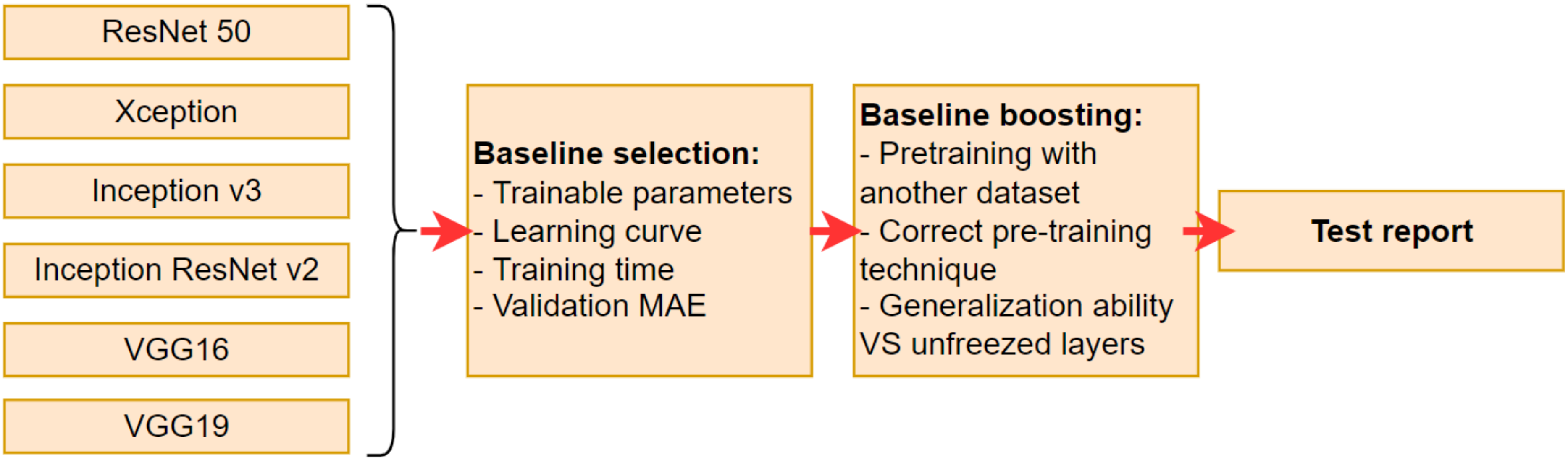
- Collected in the work of Chen et al. [1]
- **2,000** sequential images
- **480 × 640** pixels

- Trainval / Test split – **80% / 20%** (400 + 1600 images)
- Train / Val split – **80% / 20%** (1280 + 320 images)
- Equally representative subsets

[1]. Ke Chen et al. Feature mining for localised crowd counting.

Method

Pipeline



- Models are pretrained on ImageNet dataset
- Target metrics – MAE
- Optimized metrics – MSE

Selected architectures

Baseline models
ResNet 50
Xception
Inception v3
Inception ResNet v2
VGG16
VGG19

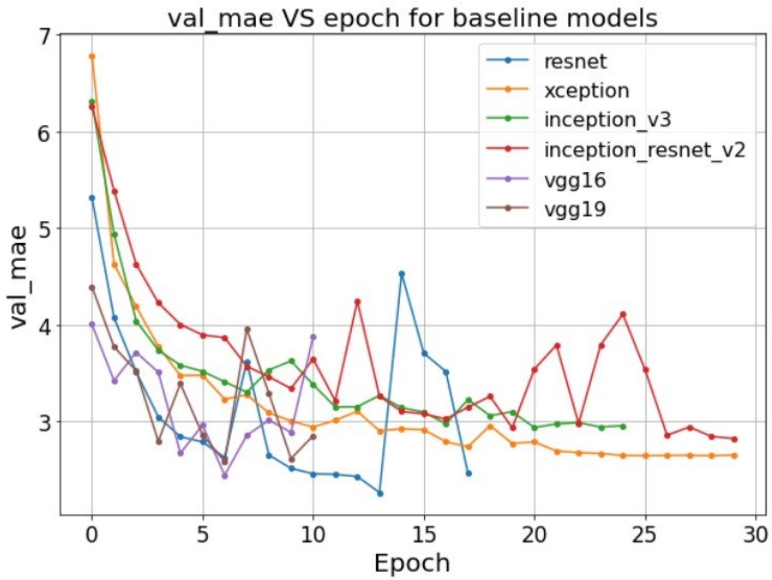
\+

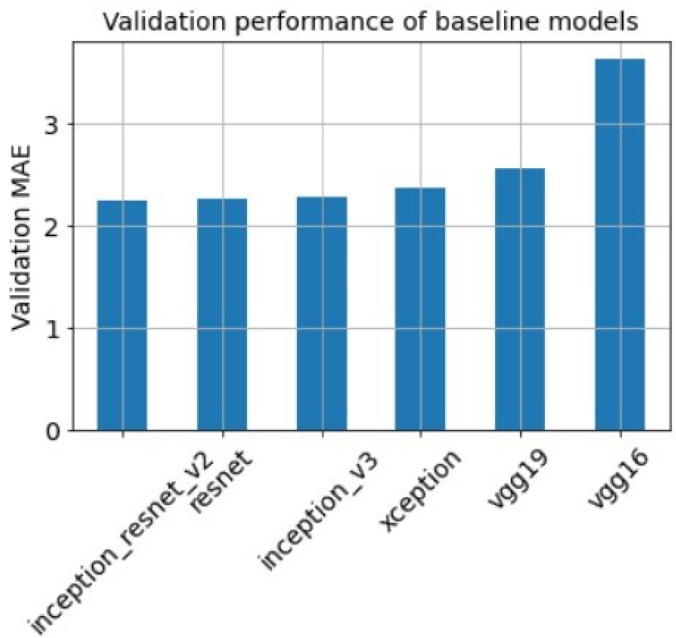
Baseline classification block
Dense (400, ReLU)
Dropout (0.3)
Dense (400, ReLU)
Dropout (0.3)
Dense (1, Linear)

- Each model:
  - Retains only feature extractor
  - Appends a classification block
  - Unfreezes 7 top layers (classification block + 2 layers of feature extractor)
- Training hyperparameters:
  - ≤ 30 epochs
  - Adam optimizer
  - Batch size of 64
  - EarlyStopping and ReduceLROnPlateau callbacks

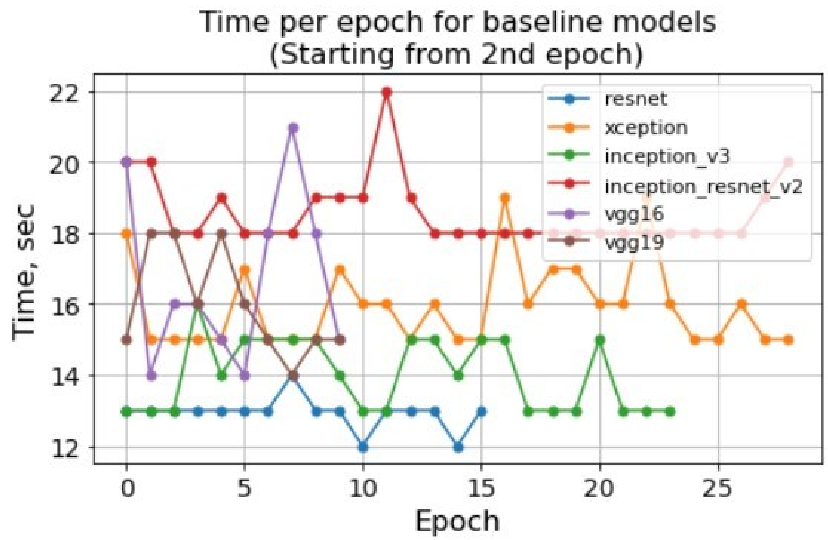
Experiments.
Baseline selection

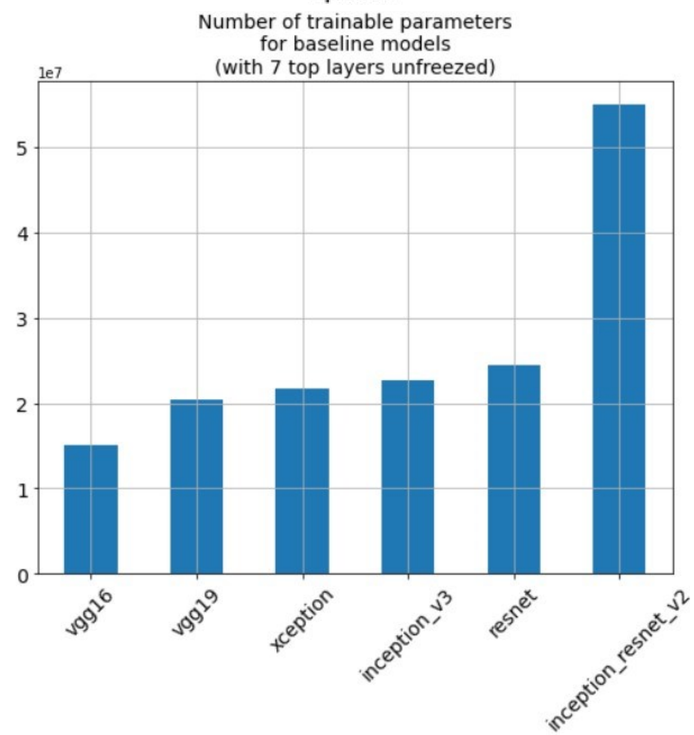
Baseline shortlisting



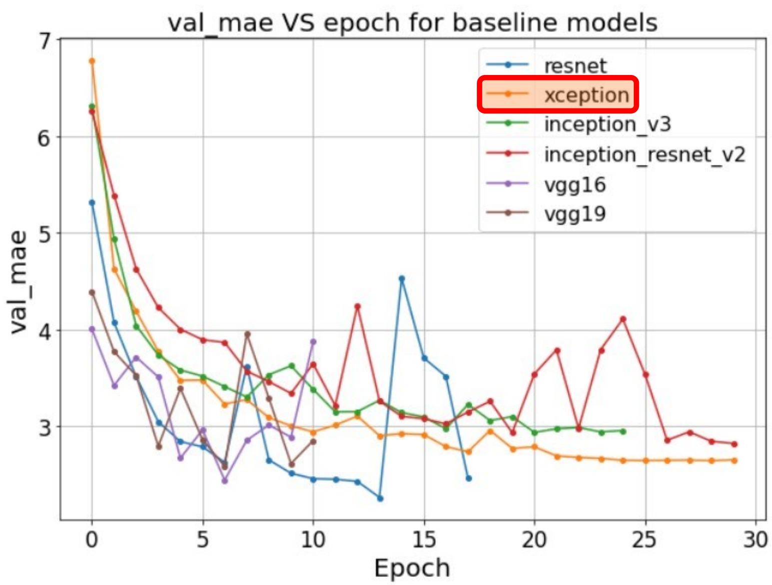


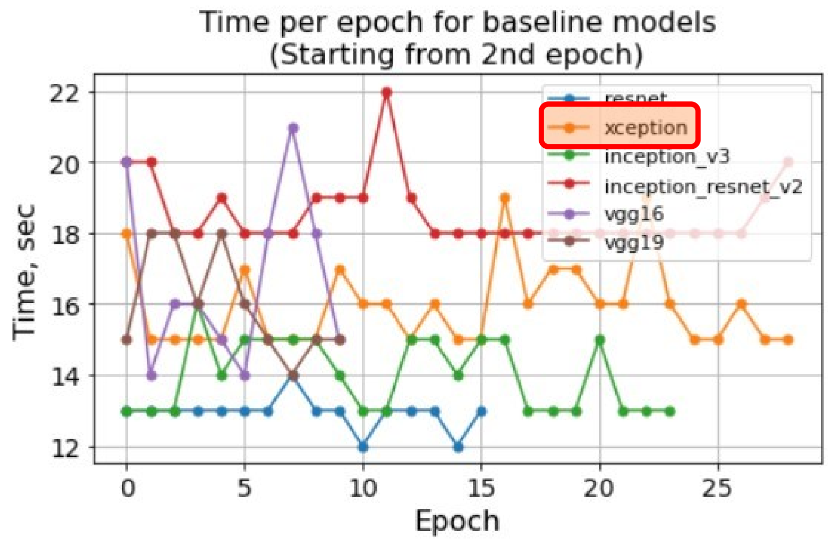
Baseline shortlisting

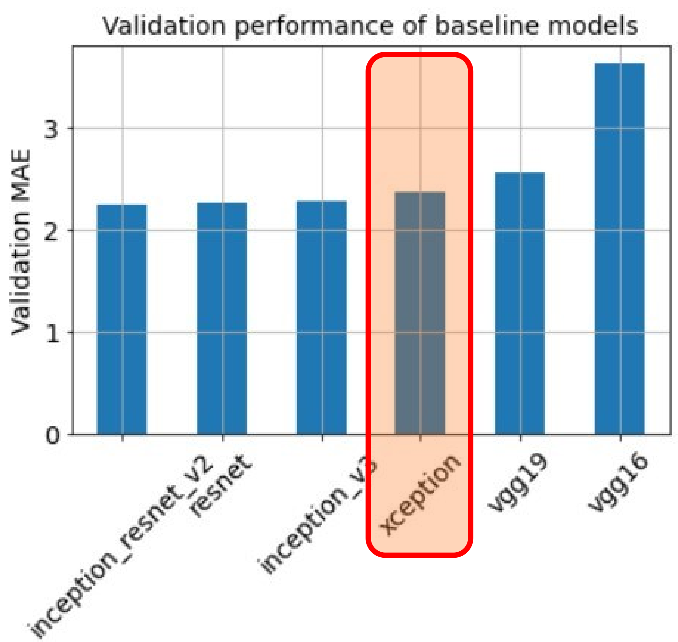


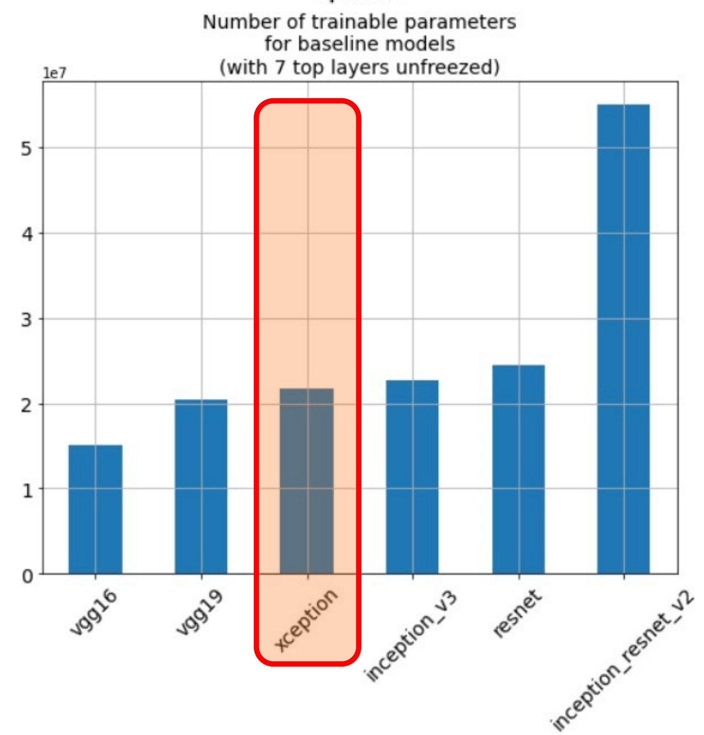


Baseline shortlisting





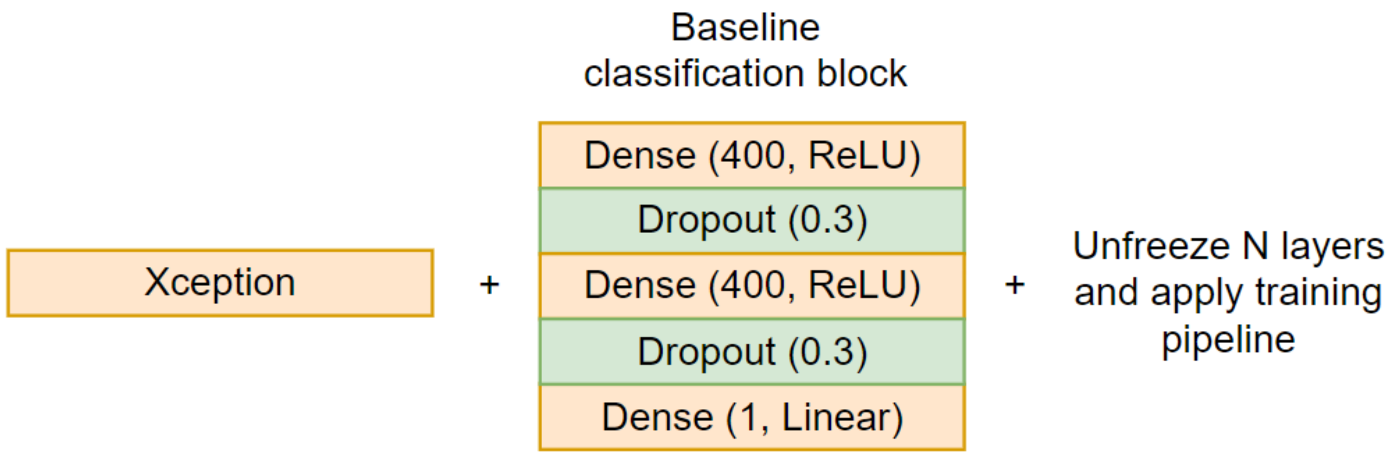


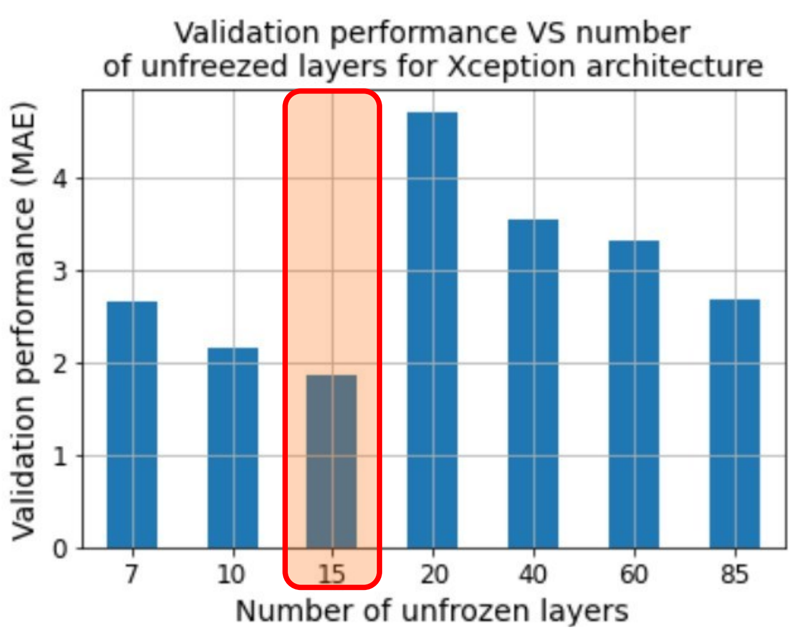


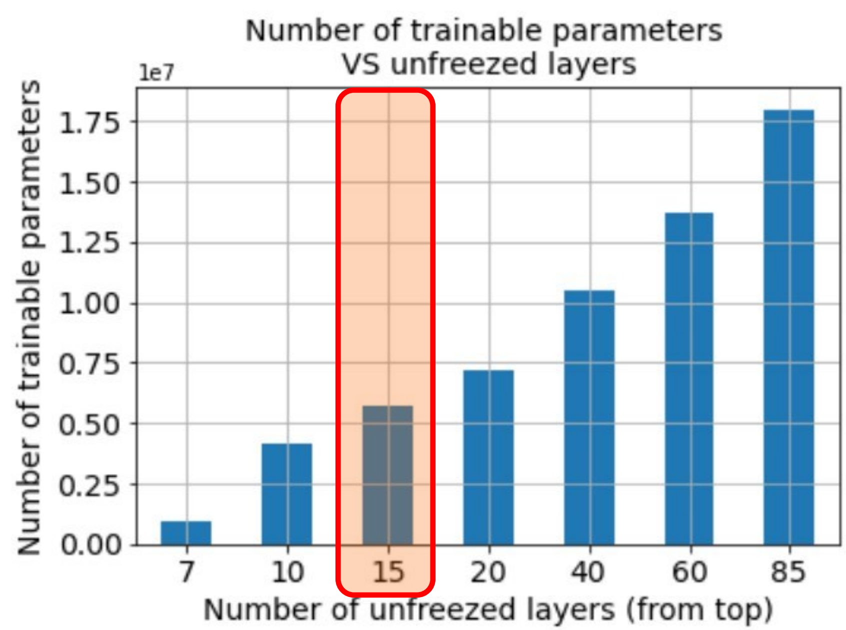
Experiments. Optimal number of unfrozen layers

The best number of unfrozen layers

- Idea for the experiment is taken from [1]







[1]. A. Geron. Hands on Machine Learning Guide

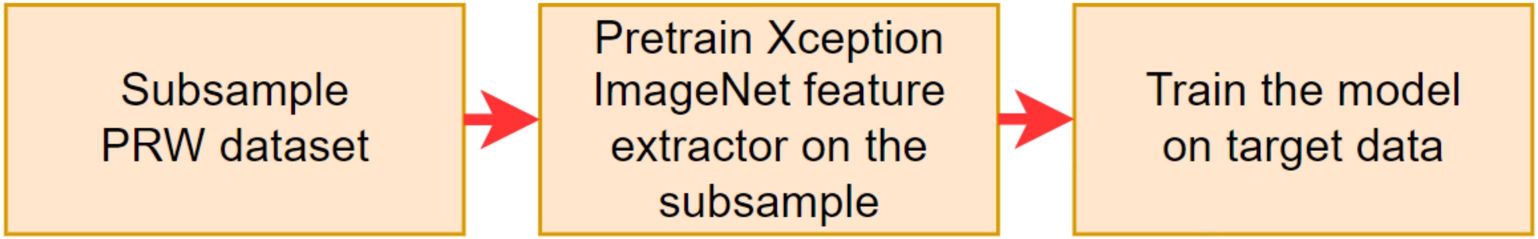
Experiments. Pretraining on another dataset

PRW dataset

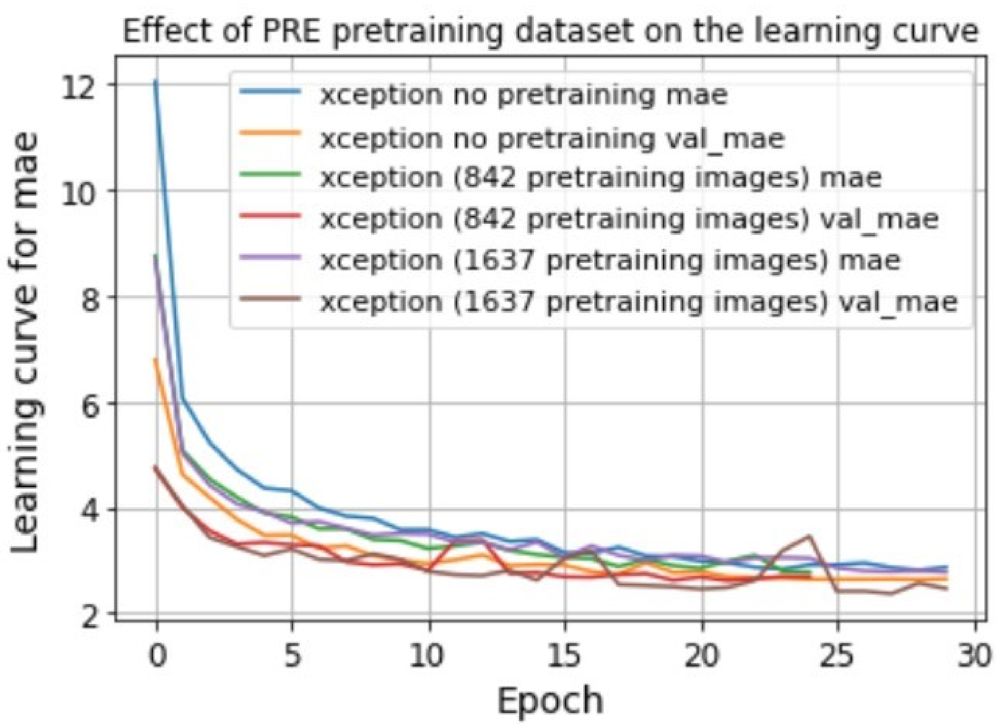
- PRW (Person Reidentification in the Wild) dataset is taken from [1]
- Dataset parameters:
  - 11,816 sequential images
  - 1080 × 1920 pixels

[1]. Liang Zheng et al. Person reidentification in the wild.

Pretraining pipeline and results

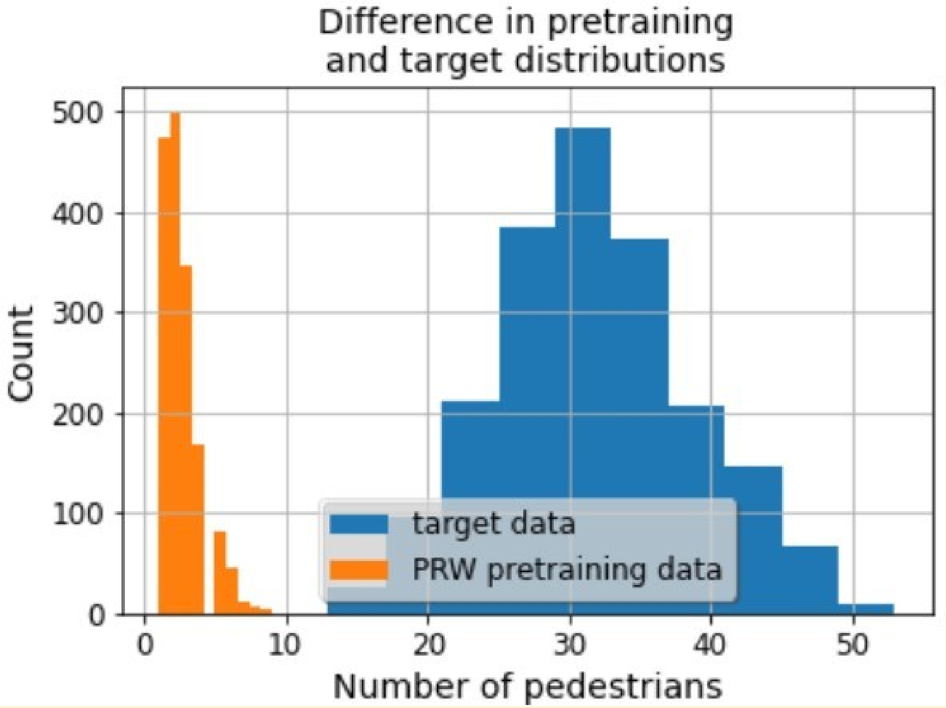


	Pre-training	Training
Number of epochs	≤20	≤30
Dataset	PRW dataset	Mall dataset
Batch size	32	
EarlyStopping	+	
Reduce LROnPlateau	+	

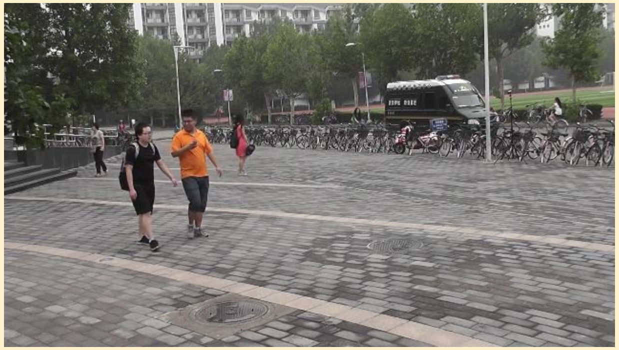


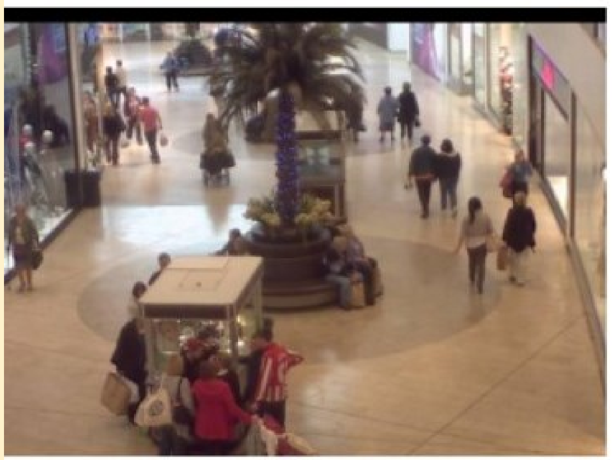
Reasons for unsuccessful pretraining

Reason1: different target distribution



Reason 2: different scale of humans, point of view and outfit

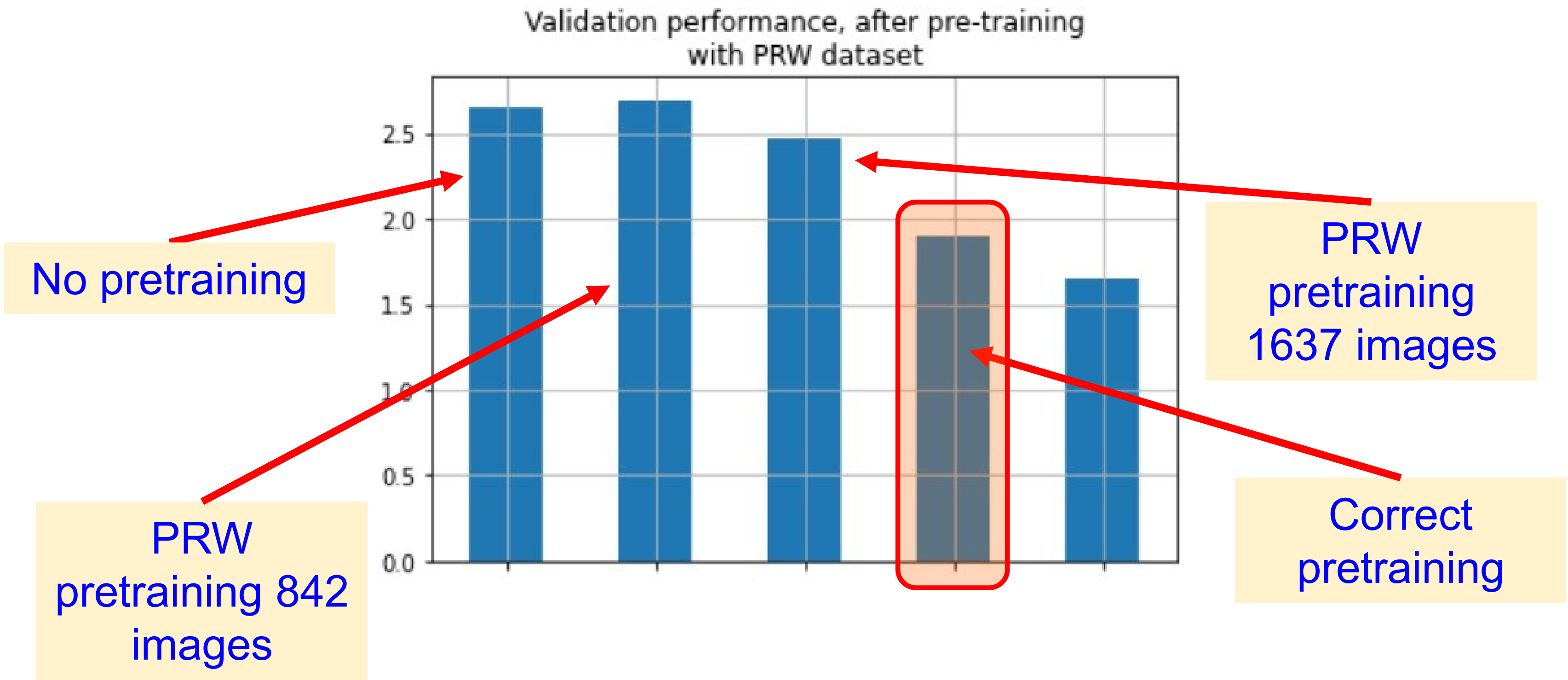




Experiments.
Correct pretraining technique

Correct pretraining

- Correct pretraining algorithm [1]:
  - Unfreeze only the layers with randomly initialized weights
  - Train for 5-6 epochs
  - Unfreeze all layers, that must be reused (number found experimentally)
  - Consider decreasing learning rate
  - Initiate training procedure



[1]. A. Geron. Hands on Machine Learning Guide

Test report

Final solution

- Final solution parameters:
  - Xception feature extractor, pretrained on ImageNet
  - Deeper classification block with ELU activations
  - Correct pretraining procedure [1]

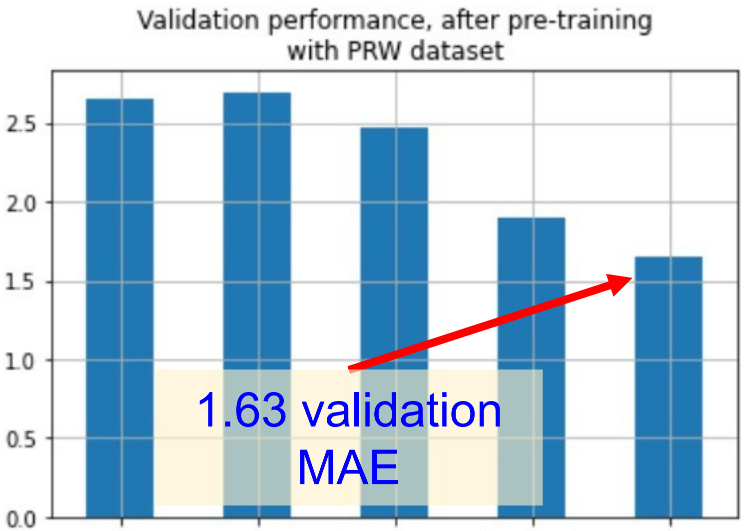
Final classification block

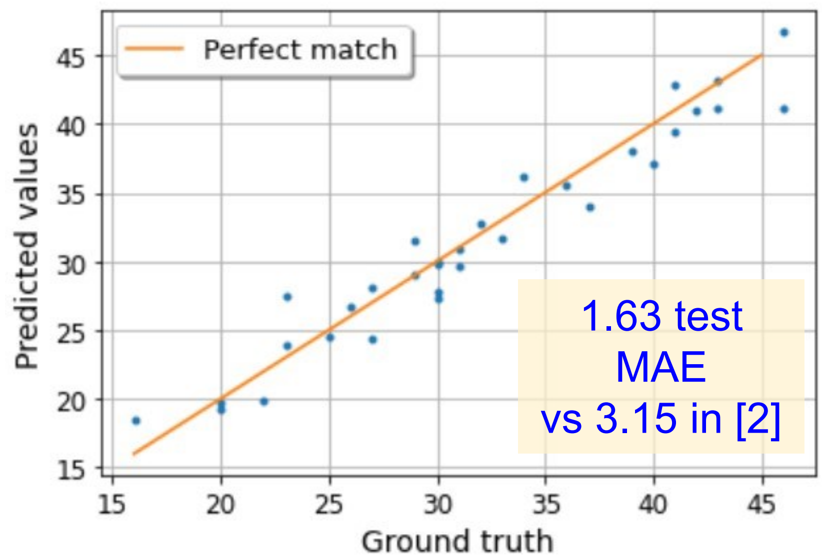
Dense (500, ELU)
Dense (400, ELU)
Dense (400, ELU)
Dropout (0.3)
Dense (400, ELU)
Dropout (0.3)
Dense (1, ReLU + he_normal)

	Pre-training	**Training**
Number of epochs	6	≤40
Number of unfreezed layers	7	17
Dataset	Mall dataset	
Optimizer	Adam	
Learning rate	0.001	0.0004
EarlyStopping	-	+
ReduceLR OnPlateau	-	+

[1]. A. Geron. Hands on Machine Learning Guide

Final solution





[1]. A. Geron. Hands on Machine Learning Guide.
[2]. Ke Chen et al. Feature mining for localised crowd counting.

Conclusions

- CNN approach is competitive with regression-based, clustering-based, and detection-based techniques:
  - No complex pipelines, **end-to-end learning**
  - **Automatic feature extraction**
- 5 / 6 architectures achieved **< 3 validation MAE**, on a baseline level:
  - ResNet 50
  - Xception
  - Inception v3
  - Inception ResNet v2
  - VGG19
- Pretraining on **PRW dataset** helped to reduce initial loss, but did not increase generalization ability
- **Correct pre-training procedure** reduced validation MAE to <2
- Final solution achieves **1.63 validation MAE, 1.64 test MAE,** which outperforms regression based approach [1] and object detection approach [3]

[1]. Ke Chen et al. Feature mining for localised crowd counting.
[2]. A. Geron. Hands on Machine Learning Guide.
[3]. https://www.kaggle.com/code/ekaterinadranitsyna/crowd-detection-model