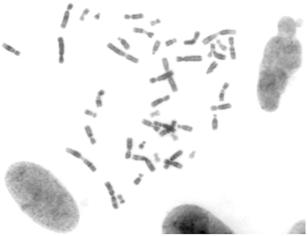
FACULTY OF ELECTRICAL
ENGINEERING
AND COMMUNICATION
department
of radio electronics

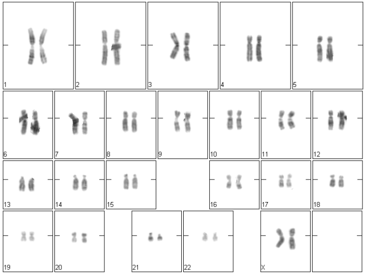
Chromosome Segmentation and Classification

Kristýna Pijáčková

Karyotyping

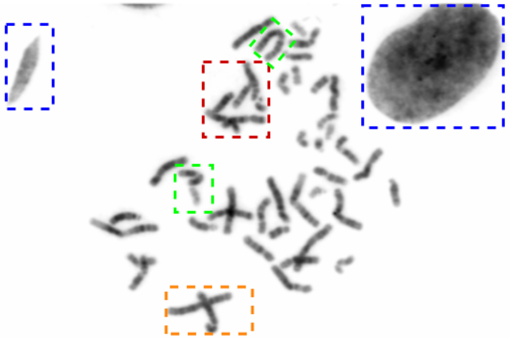
- Technique used for analyzing chromosomal structures
- Chromosomes are separated, organized and inspected for any abnormalities
- Abnormalities may lead to diagnosing specific birth defects, genetic disorders or cancer





Challenges of chromosome segmentation

- Laborious and time-consuming
- Demands high level of expertise
- Overlapping chromosomes
- Touching chromosomes
- Bent chromosomes
- Non-chromosomal objects

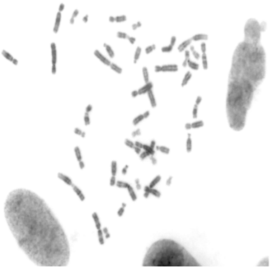


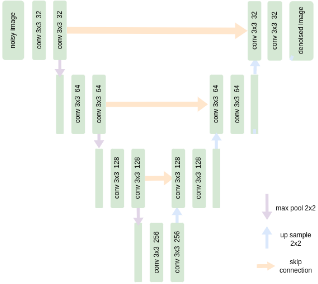
Segmentation pipeline

- Image denoising
- Chromosome location and cluster identification
- Instance segmentation with Mask R-CNN
- Test-Time Augmentation

Image denoising

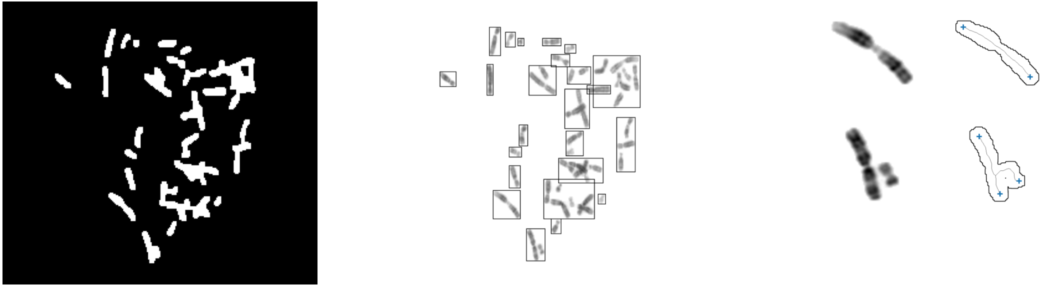
- U-Net architecture to get rid of nucleus cells and dust particles
- Trained and tested on BioImlab dataset from University of Padova





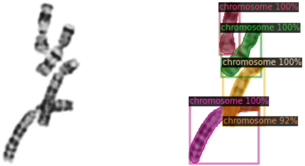
Chromosome location and cluster identification

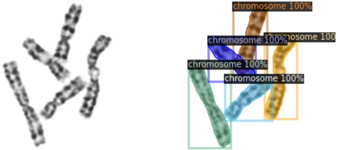
- Binary mask obtained with Otsu's thresholding
- Detection and extraction of chromosomes with OpenCV bounding boxes
- Sorting of single and overlapping/touching chromosomes



Instance segmentation

- Clinical dataset from Guangdong Women and Children's Hospital
- Mask R-CNN architecture with ResNet50 as backbone
- Flipping augmentation during training





Test-time augmentation

- Increases output precision of segmented masks
- Input image is flipped by 90°, 180°, and 270° and segmented
- Predicted masks are matched by choosing mot probable fit with IoU over 0.5
- Output masks are product of averaging the matched augmented masks

Examples of segmented chromosomes

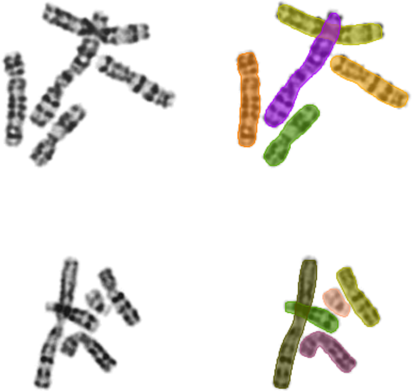


Fig. 1: Correct segmentation

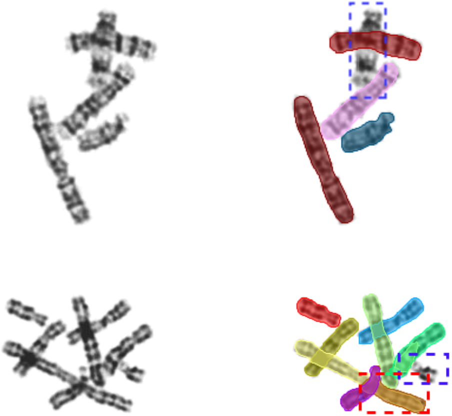
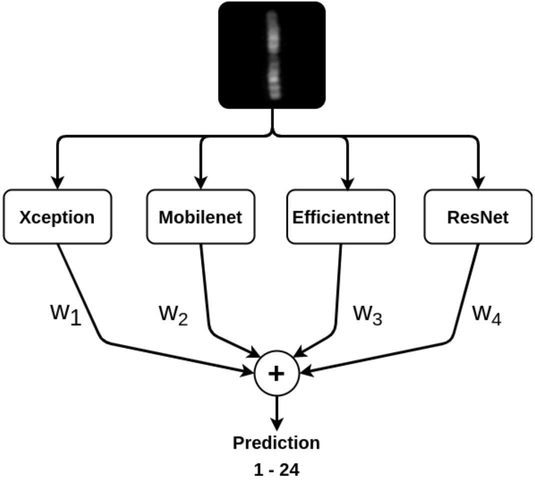


Fig. 2: Incorrect segmentation

Model Ensembling for Chromosome Classification

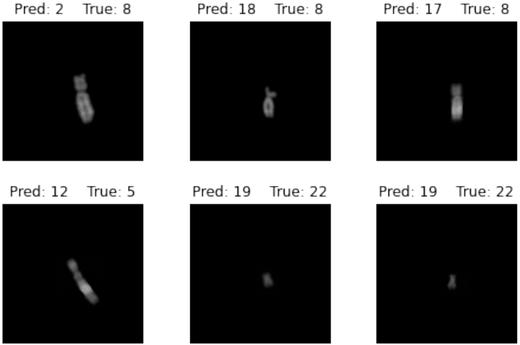
- Improves classification results by pooling multiple models together
- Based on assumption that different, good models can extract different features
- Used models - Xception, MobilenetV2, EfficientnetB2, ResNet18



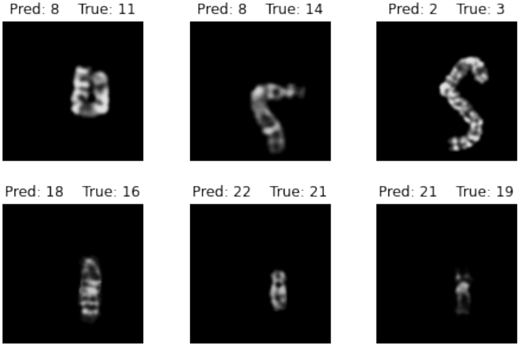
Single model vs. Model ensembling

Model	**Dataset**	**Accuracy**	**Dataset**	**Accuracy**
Xception	BioImlab	91.92 %	CIR	96.56 %
MobilenetV2		83.46 %		90.14 %
EfficientnetB2		92.29 %		95.41 %
ResNet18		84.45 %		94.50 %
Model Ensemble		**94.78 %**		**97.48 %**

Examples of misclassifications



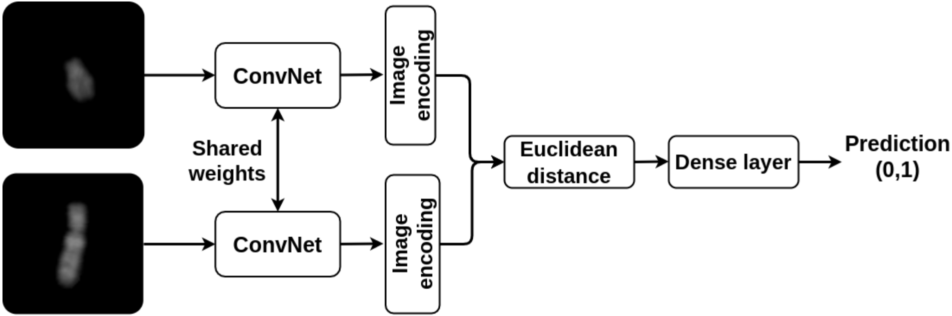
Bioimlab dataset



CIR dataset

Siamese network

- Can compare similarity between two images
- Top-3 accuracy ∼98 % for both datasets
- Incorrectly paired chromosomes are checked and rearranged based on the similarity of the pairs



Thank you for your attention!