Removing Background from Portrait Images using U^2-Net trained with Cyclical Learning

Presented By:-
Rajdeep Dutta
Gurjot Singh

YORK UNIVERSITÉ UNIVERSITY

OUTLINE

1. Project Aim
2. Data Description
3. Implementation Details
    a. Image Segmentation
    b. Image Matting
    c. U^2Net
    d. Cyclical Learning
    e. Edge Mask / Boundary based loss
4. Training Environment
5. Intermediate Results
6. Future Work

YORK UNIVERSITÉ UNIVERSITY

PROJECT AIM

- Remove Background from Human Portrait Images.
- User Image Resolution Agnostic.
- Use U^2-Net[1] model architecture.
- Cyclical Learning[2]
- Akin to Image Matting[3] (different from Image Segmentation).
- Use Boundary based weighted RMSE loss, based on the MODNet paper[4]

[1]: Qin, Xuebin & Zhang, Zichen & Huang, Chenyang & Dehghan, Masood & Zaïane, Osmar & Jagersand, Martin. (2020). U2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognition. 106. 107404. 10.1016/j.patcog.2020.107404.

[2]: L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 2017, pp. 464-472, doi: 10.1109/WACV.2017.58.

[3]: Li, J., Zhang, J., & Tao, D. (2023). Deep Image Matting: A Comprehensive Survey. *ArXiv, abs/2304.04672*.

[4]: Ke, Zhanghan & Sun, Jiayu & Li, Kaican & Yan, Qiong & Lau, Rynson. (2022). MODNet: Real-Time Trimap-Free Portrait Matting via Objective Decomposition. Proceedings of the AAAI Conference on Artificial Intelligence. 36. 1140-1147. 10.1609/aaai.v36i1.19999.

YORK UNIVERSITÉ UNIVERSITY

PROJECT AIM

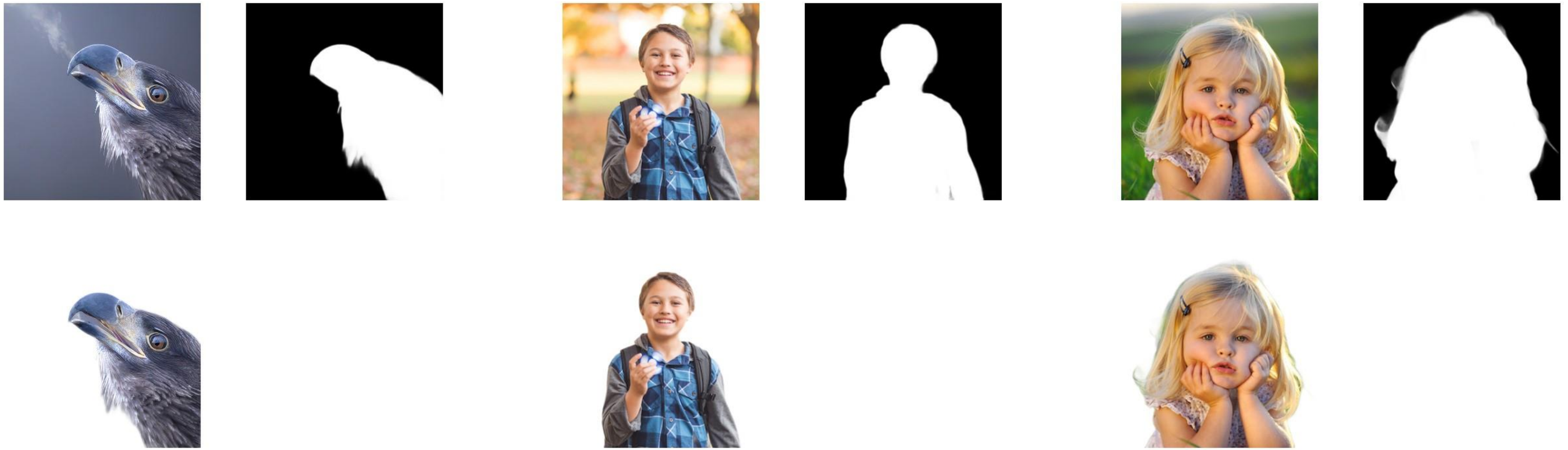


Figure: Background removal.

DATA DESCRIPTION

- Dataset: EasyPortrait[1]
- Number of Images: 40,000
- ~38.3K FullHD+ images
- Size of dataset: 91.78GB
- Number of Unique Persons: 13,705

[1]: https://github.com/hukenovs/easyportrait

YORK UNIVERSITÉ UNIVERSITY

DATA DESCRIPTION

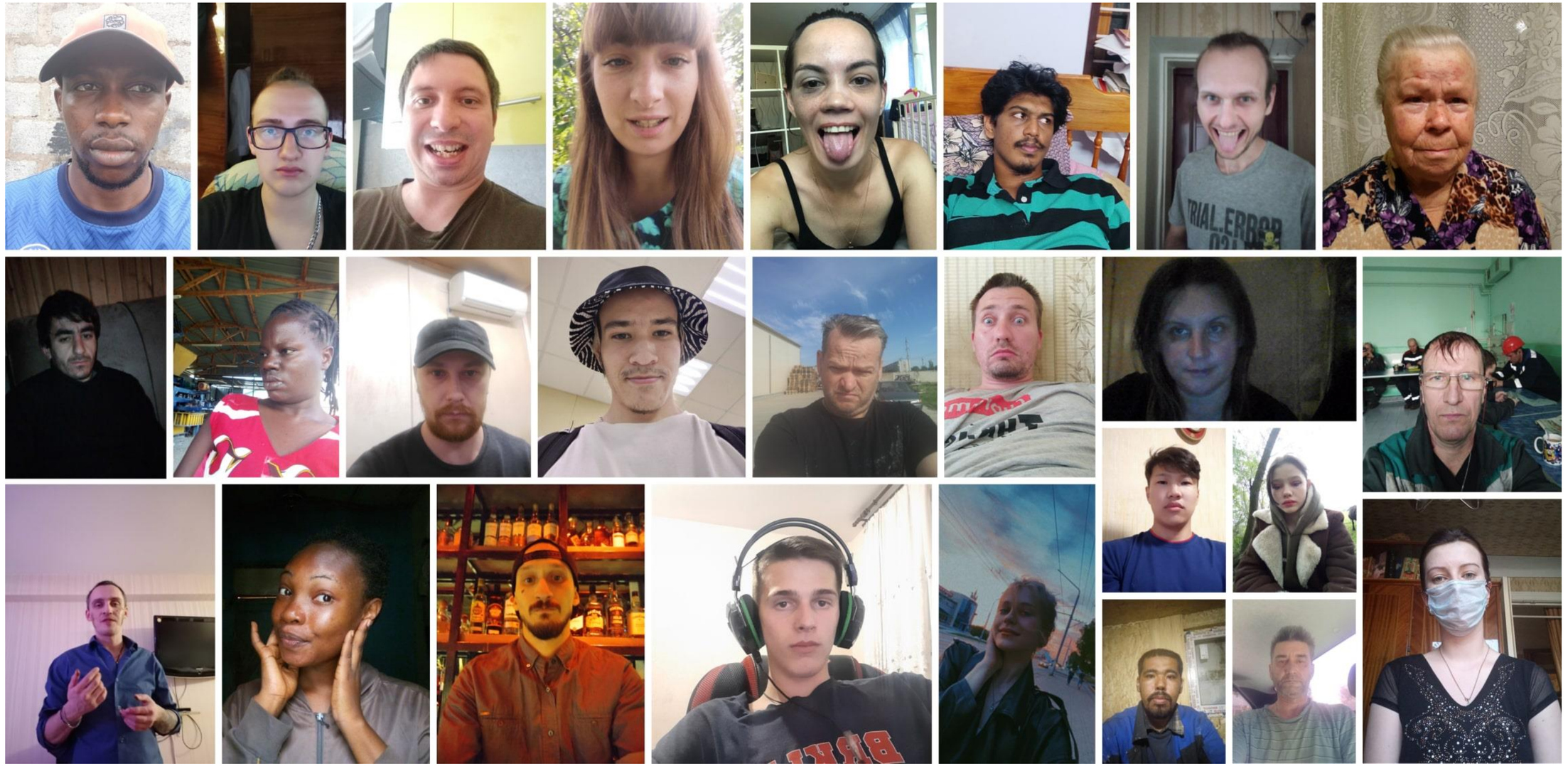


Figure: Example of Images in the EasyPotrait dataset

DATA DESCRIPTION

Figure: Each part of the face are mapped to different pixels for the task of segmentation

IMPLEMENTATION DETAILS

Image Segmentation

- Assigns each pixel to a discrete label.
- Output is a binary mask delineating foreground objects
- Each pixel is classified to one label
- Common techniques include thresholding, region-based segmentation, edge detection.

Figure: Image Segmentation

YORK UNIVERSITÉ UNIVERSITY

IMPLEMENTATION DETAILS

Image Matting

- Estimates the opacity / alpha value of each pixel.
- Basically, Alpha Channel prediction.
- Alpha indicates coverage of foreground object vs background at that pixel.
- Output is continuous alpha matte with alpha values between 0-1 per pixel.
- Retains soft edges for semi-transparent regions like hair.

Figure: Image Matting

YORK UNIVERSITÉ UNIVERSITY

IMPLEMENTATION DETAILS

U^2Net

- Deep learning architecture for salient object detection.
- Capture more contextual information from different scales due to the mixture of receptive fields of different sizes in the proposed ReSidual U-blocks (RSU).
- Two-level nested U-structure architecture that allows the network to go deeper, attain high resolution, without significantly increasing the memory and computation cost.
- It increases the depth of the whole architecture without significantly increasing the computational cost because of the pooling operations used in these RSU blocks.

YORK UNIVERSITÉ UNIVERSITY

IMPLEMENTATION DETAILS

U²Net

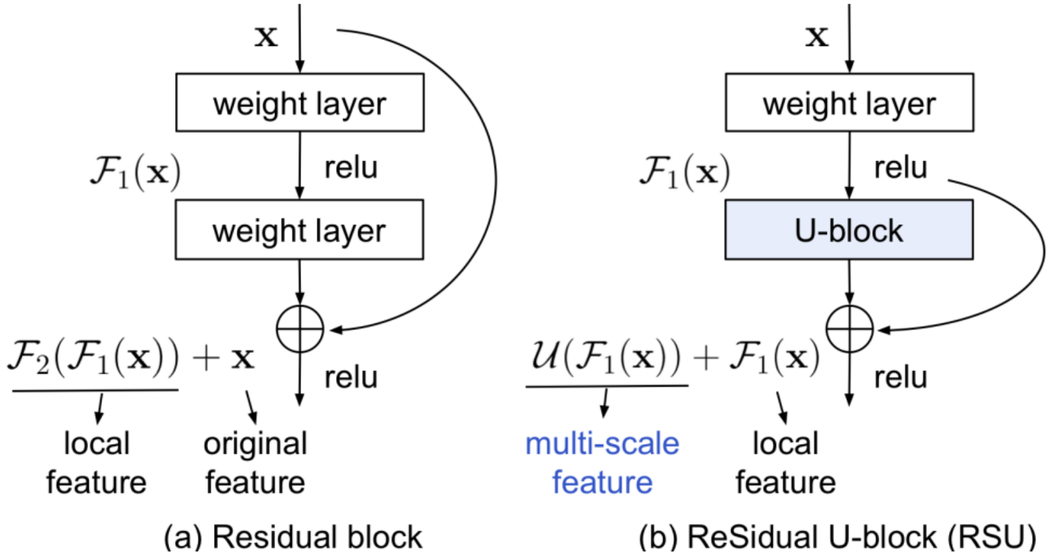


Figure: Residual block vs Residual U-block

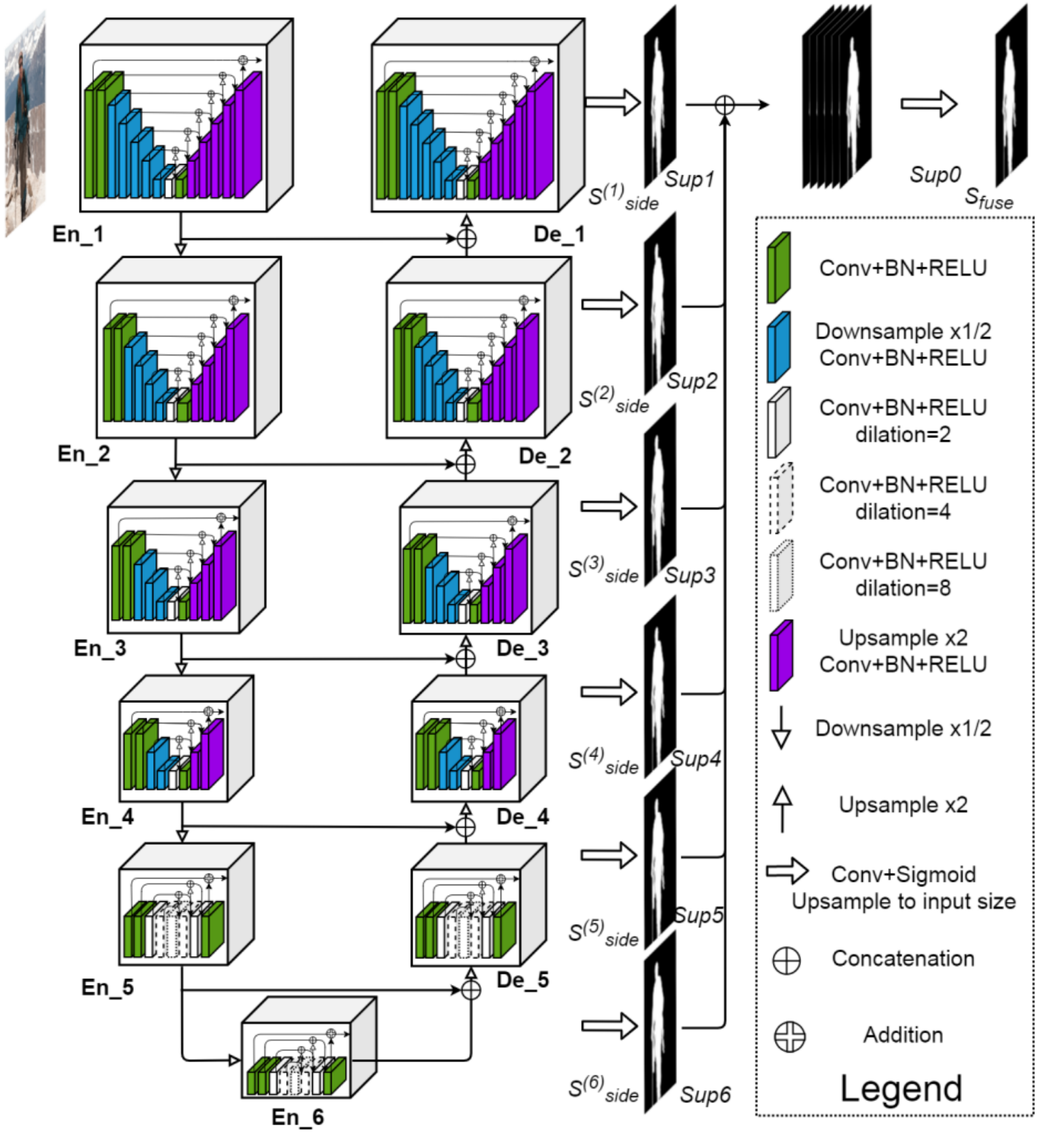


Figure: U²Net Architecture

IMPLEMENTATION DETAILS

Cyclical Learning

- Method for setting the LR - eliminates the need to experimentally find the best values and schedule for the global learning rates.
- Instead of monotonically decreasing the LR - the learning rate cyclically vary between reasonable boundary values. '*Annihilate*' at the end.
- The oscillation can be linear, triangular, or any other shape.
- Achieves improved classification accuracy without a need to tune and often i

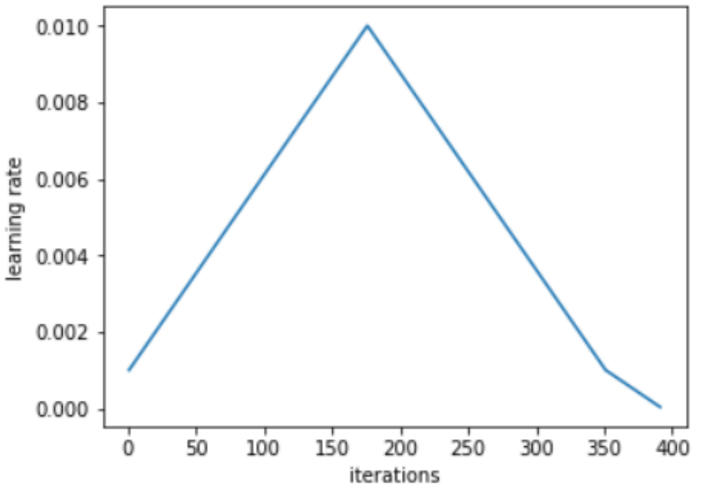


Figure: One cycle. Annihilate at the end.

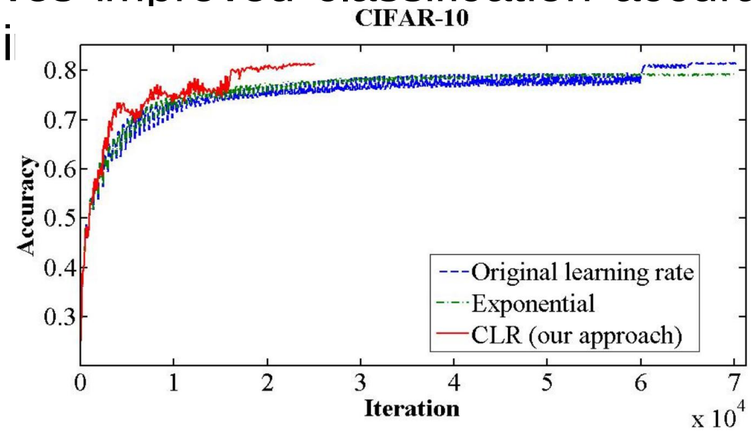


Figure: Comparison of different learning rate approaches.

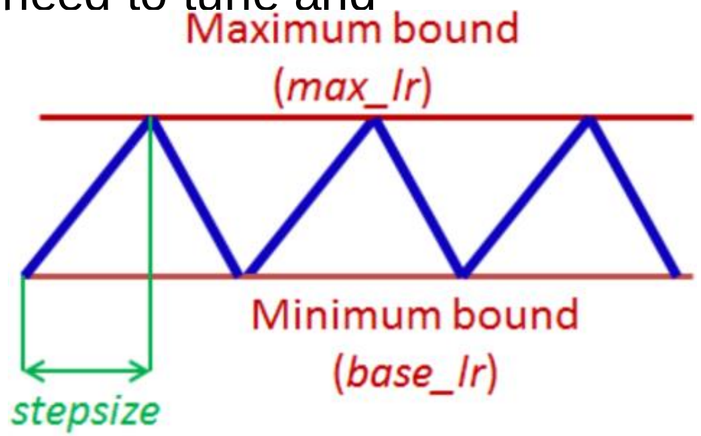


Figure: Triangular Learning rate policy. Blue line denotes LR.

IMPLEMENTATION DETAILS

Edge Mask / Boundary Based Loss

• Generate a *Transition* region around the edge or boundary of the mask.

• Use Erosion followed by Dilation operations on the ground truth mask to generate this border mask.

• Transition Mask: Pixel = 1 if it falls within the border region else 0.

• Used to focus more on the boundaries of the object where the foreground pixels slowly changes into background pixels.

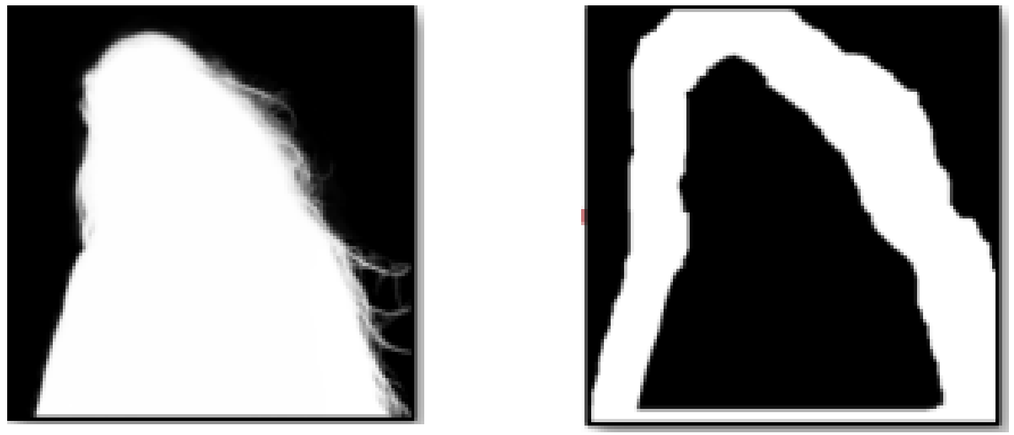


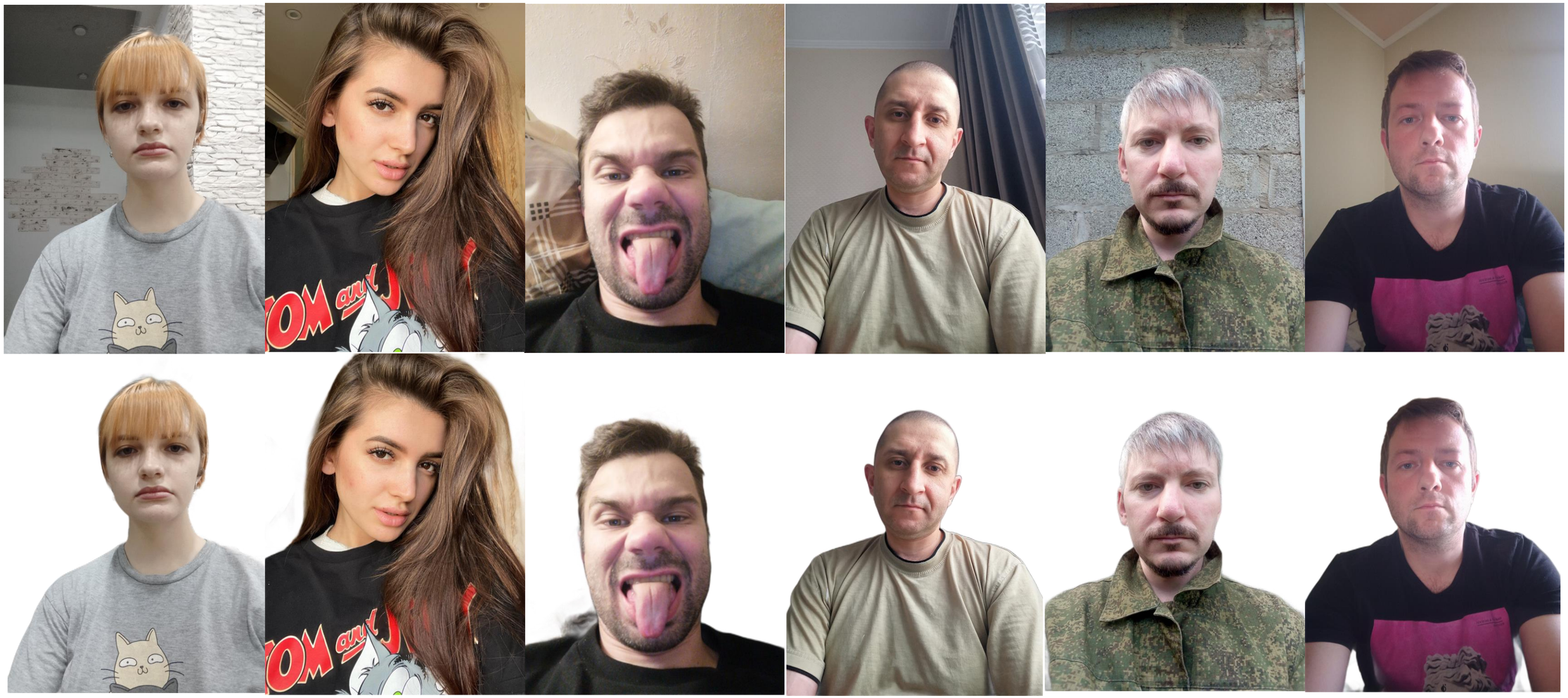
Figure: Boundary Mask in MODNet architecture

TRAINING ENVIRONMENT

- Packages / Frameworks / Compilers:
  - Python: 3.6.13
  - PyTorch: 1.14.x
  - CUDA Toolkit: 10.0.130
  - CUDNN: 7.6.5
  - NVCC: 11.7
  - Visual Studio Community 2019
- Hardware:
  - GPU: RTX 2060 Super (8GB VRAM)
  - CPU: Ryzen 3600
  - RAM: 16GB 3200MHz
- Operating System: Windows 10 Pro 64bit (v22H2)

YORK UNIVERSITÉ UNIVERSITY

INTERMEDIATE RESULTS



INTERMEDIATE RESULTS

FUTURE WORK

- Explore other architectures such as K-Net[1]
- Use Mixed Precision Training – e.g. AMP module[2] in Pytorch for QAT (Quantization Aware Training)
    - Faster Training for the same input size.
    - Increase Batch Size.
    - Lowers VRAM usage, so input size can be further increased (e.g. 720x720) for better results.
- Use Learning Rate Finder[3][4] to find the optimal max_lr in cyclical learning.

[1]: Zhang, W., Pang, J., Chen, K., & Loy, C.C. (2021). K-Net: Towards Unified Image Segmentation. *ArXiv, abs/2106.14855*.

[2]: Automatic Mixed Precision package - torch.amp — PyTorch 2.1 documentation, https://pytorch.org/docs/stable/amp.html

[3]: davidtvs/pytorch-lr-finder: A learning rate range test implementation in PyTorch, https://github.com/davidtvs/pytorch-lr-finder

[4]: Learning Rate Finder — PyTorch Lightning 1.5.10 documentation, https://lightning.ai/docs/pytorch/1.5.10/advanced/lr_finder.html

YORK UNIVERSITÉ UNIVERSITY

Thanks!