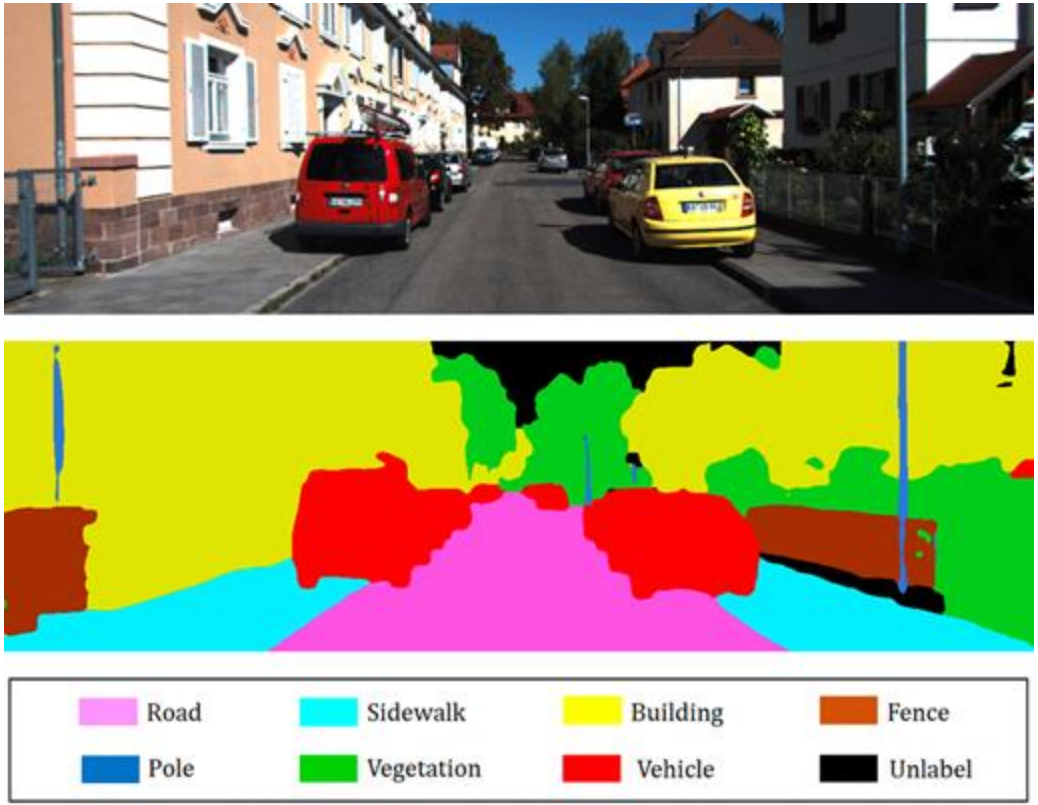
Image to image translation based on Pix2Pix GAN

Процик Олексій

Semantic segmentation



Data

Coco-stuff: 164.000 images

171 classes

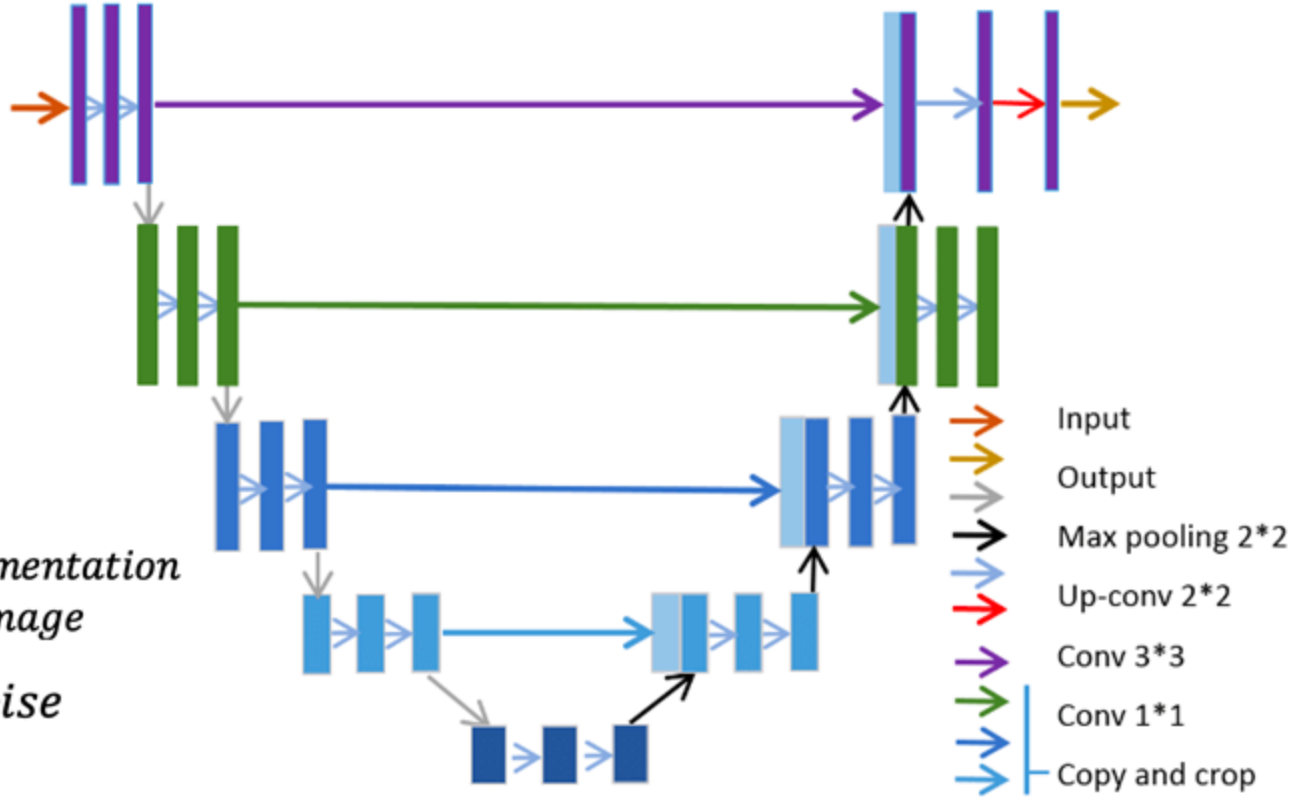
Generator

$$G_\theta(x, z) = y$$

$x \in \mathbb{R}^{b\times1\times256\times256} - semantic\ segmentation$

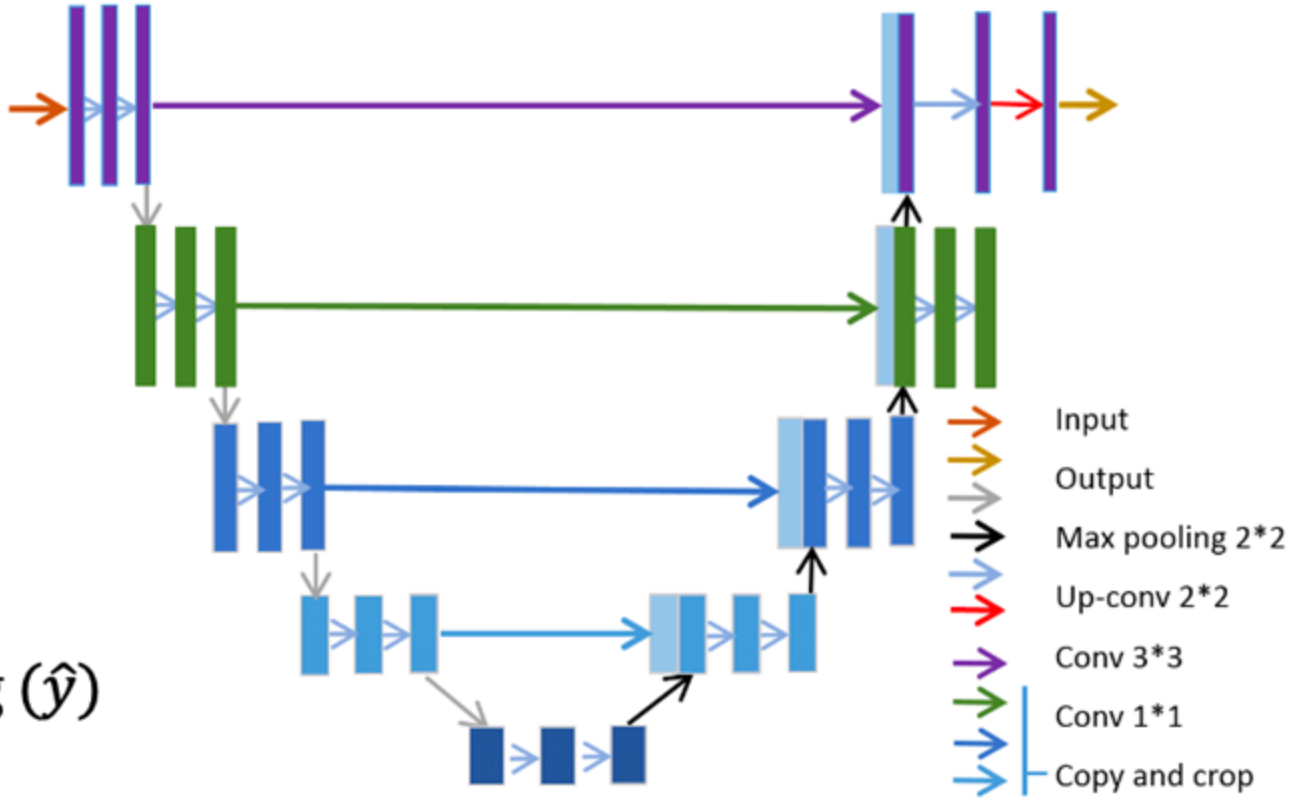
$y \in \mathbb{R}^{b\times3\times256\times256} - target\ image$

$z \sim N_{b\times256\times256}(\mu, \Sigma) - noise$



Discriminator

$$D_\phi(y) = x$$

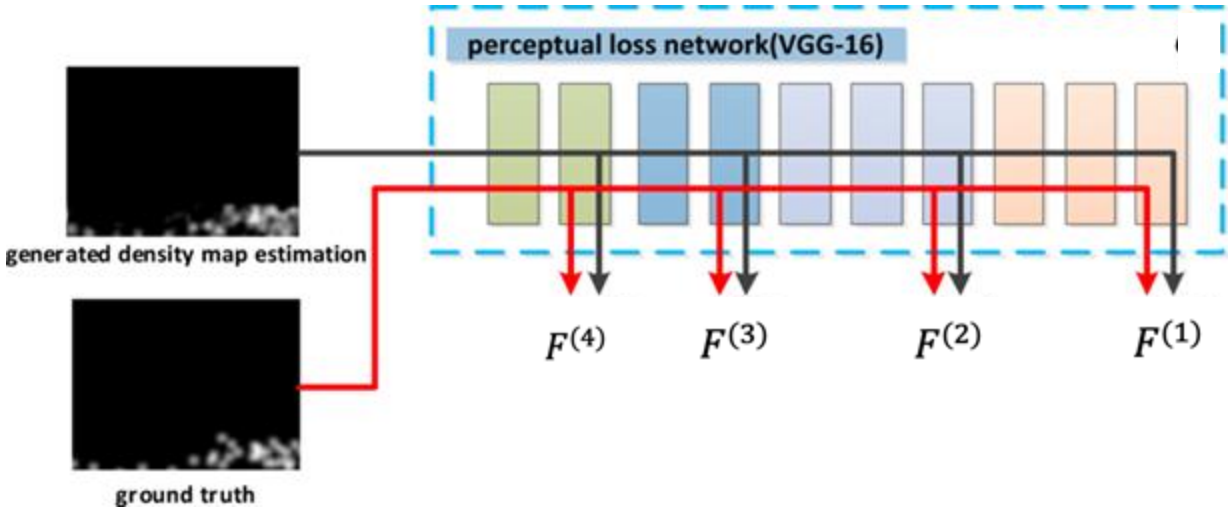


$$\mathcal{L}_D(y, \hat{y}) = -\sum_{i=0}^{172} y_i \log(\hat{y})$$

VGG Loss

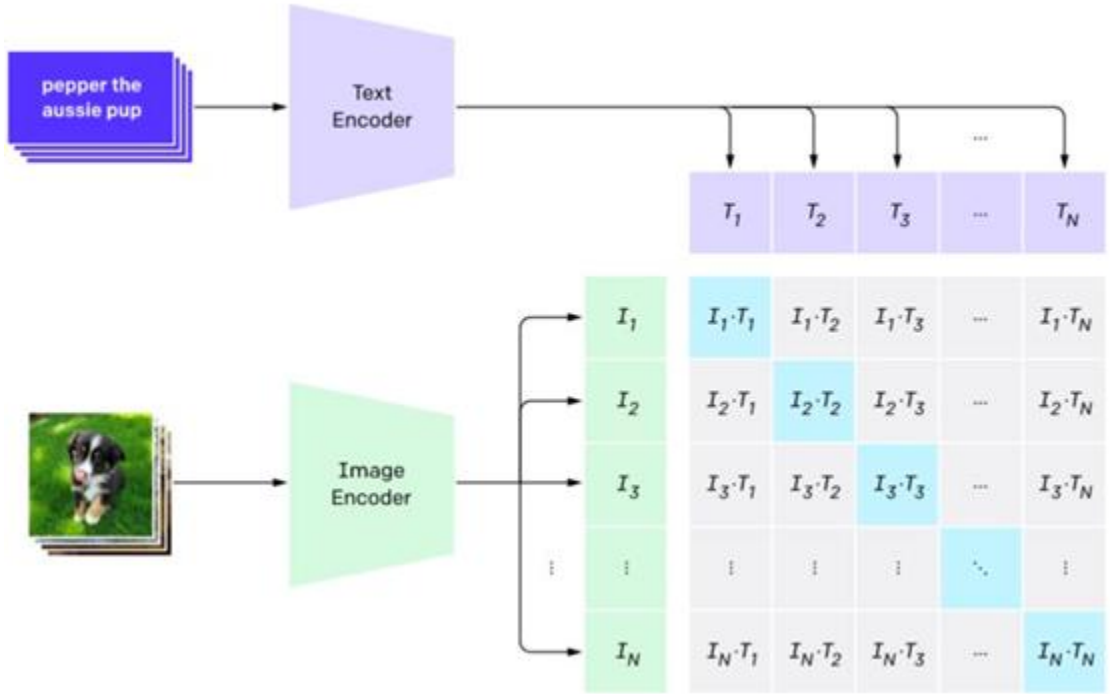
$$\mathcal{L}_{VGG}(y, \hat{y}) = \sum_{i=1}^{4} M_i[\left|F^{(i)}(y) - F^{(i)}(\hat{y})\right|_1]$$

$$\lambda_{VGG} = 10, N = 5, M_1 = \frac{1}{32}, M_2 = \frac{1}{16}, M_3 = \frac{1}{8}, M_4 = \frac{1}{4}, M_5 = 1$$



Clip Loss

$$\mathcal{L}_{CLIP}(y, \hat{y}) = \frac{y \cdot \hat{y}}{\|y\| \, \|\hat{y}\|} = \frac{\sum_i y_i * \hat{y}_i}{\sqrt{\sum_i y_i^2} + \sqrt{\sum_i \hat{y}_i^2}}$$



Generative Adversarial Network

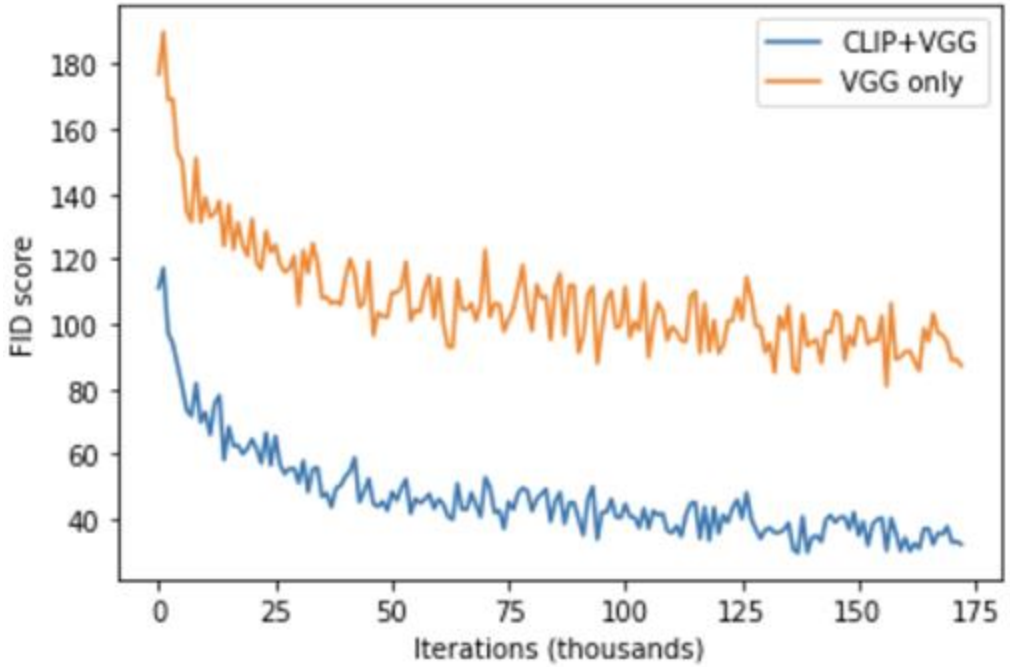
$$G^* = \min_{\theta} \max_{\phi} \lambda_{VGG} * \mathcal{L}_{VGG}\big(y, G(x,z)\big) + \lambda_{CLIP} * \mathcal{L}_{CLIP}\big(y, G(x,z)\big) + \lambda_D * \mathcal{L}_D\big(y, G(x,z)\big)$$

$$D^* = \min_{\phi} \mathcal{L}_D(y, G(x,z))$$

$$\lambda_{VGG} = 10, \lambda_{CLIP} = 30, \lambda_D = 2.5$$

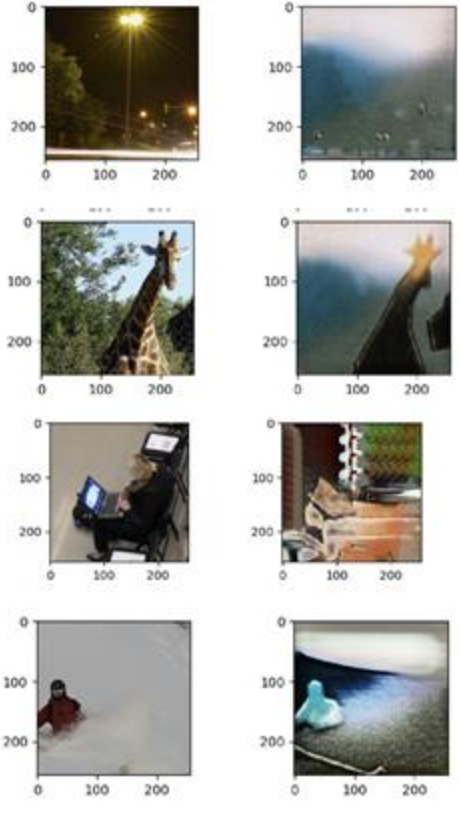
Results

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2})$$

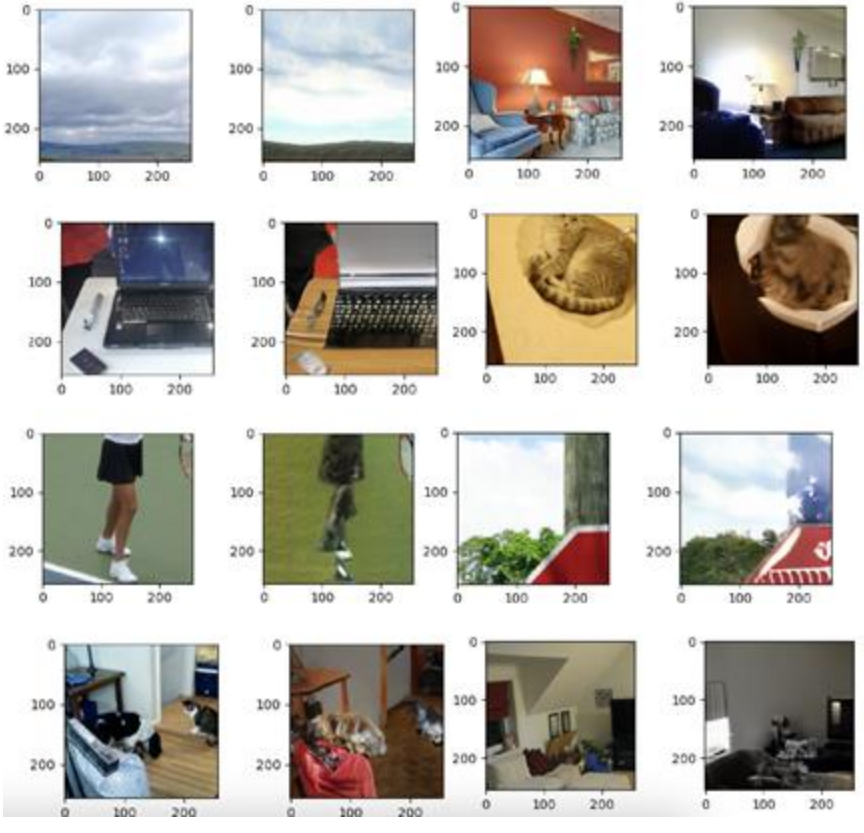


Results

VGG



VGG+CLIP



Thank you for your attention!