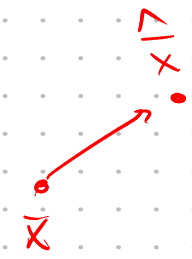
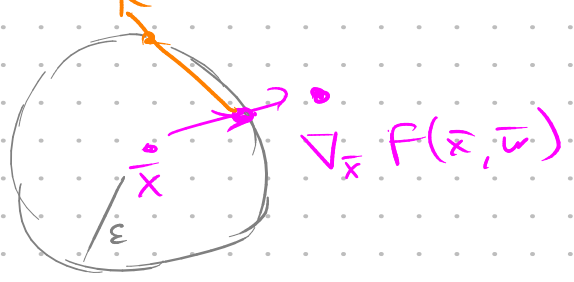


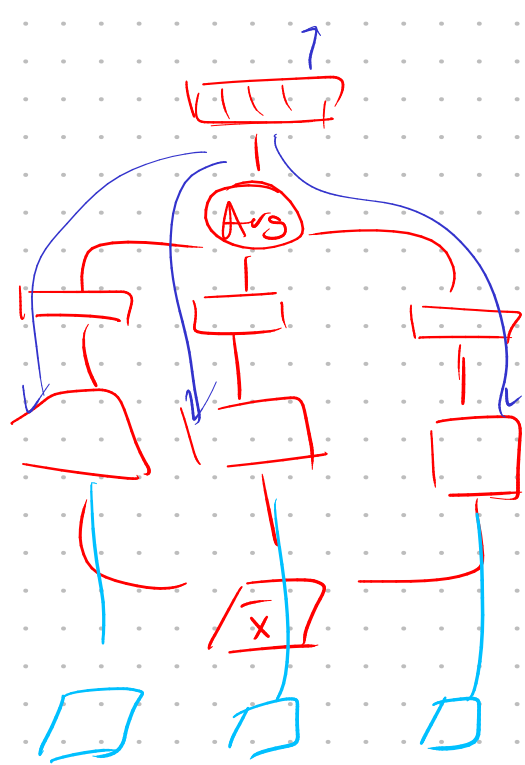
$$\nabla_{\bar{w}} F(\bar{x}, y, \bar{w})$$

$$\nabla_{\bar{x}} F(\bar{x}, y, \bar{w})$$



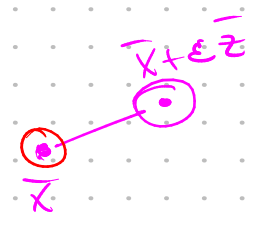
$$\bar{w}^T \hat{\bar{x}} = \bar{w}^T \bar{x} + \bar{w}^T \bar{z}$$

bag of freebies
YOLOv4



① Augmentations

②



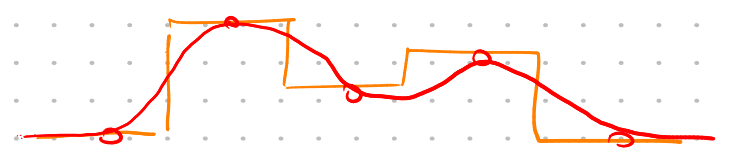
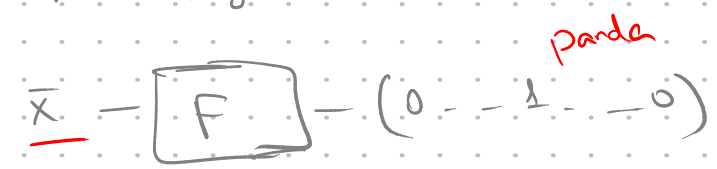
self-adversarial training

$$\hat{L}(\bar{w}, \bar{x}, y) = \alpha L(\bar{w}, \bar{x}, y)$$

$$+ (1-\alpha) L(\bar{w}, \bar{x} - \epsilon \nabla_{\bar{x}} L_y(\bar{w}, \bar{x}, y), y)$$

- White-box attacks
- Gray-box attacks
- Black-box attacks

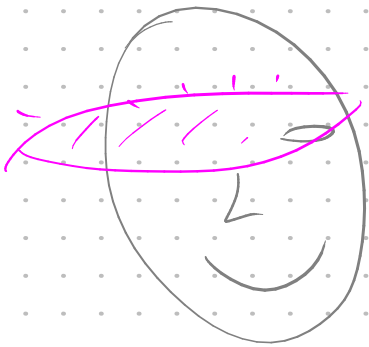
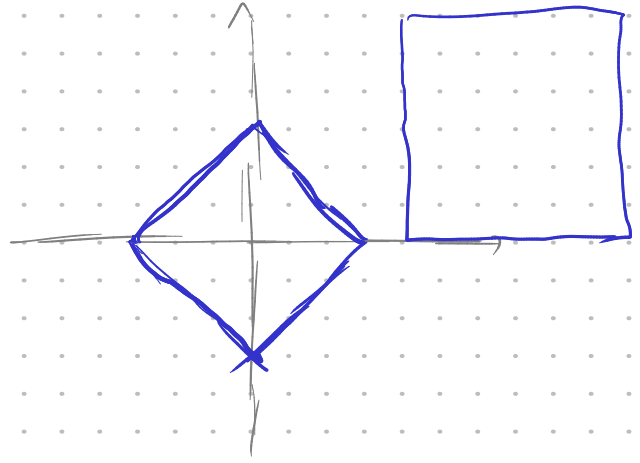
masked gradient



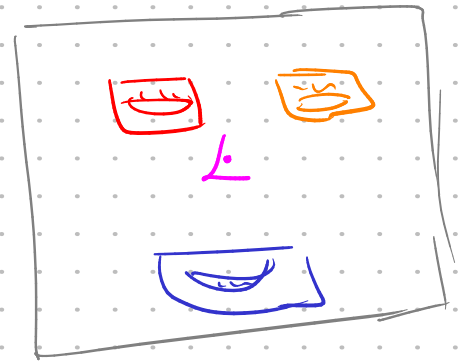
Capsule networks

\mathcal{D}

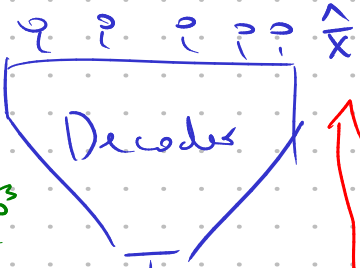
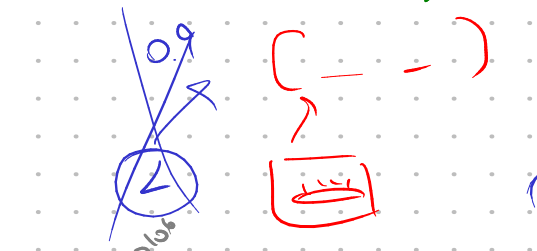
\mathcal{R}



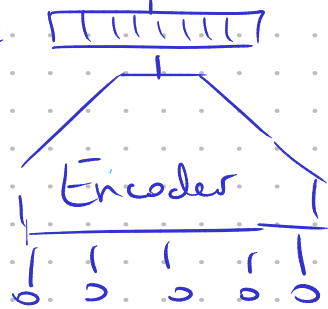
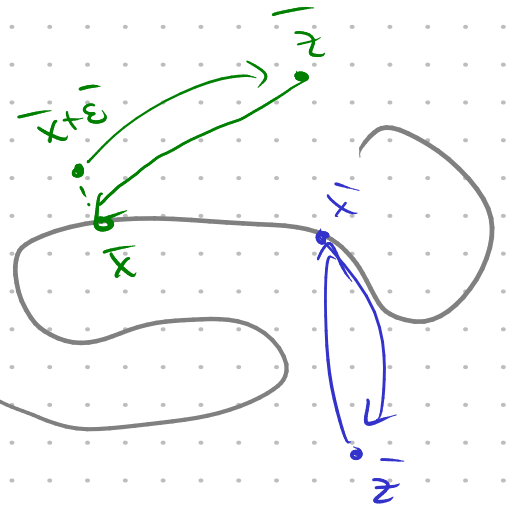
face



[conf, x, y, w, h]



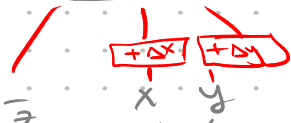
$\|x - \hat{x}\| \rightarrow \min$



denoising autoencoder

$x + E$
($x + \Delta x, y + \Delta y$)

Dec



Enc

