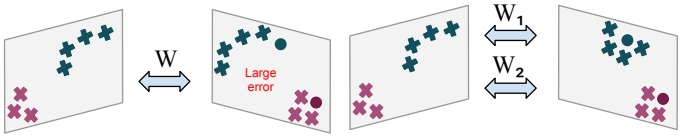
Multimodal Deep Learning

Zeynep Akata

Zero-Shot Learning

Latent Embeddings for Zero-Shot Image Classification

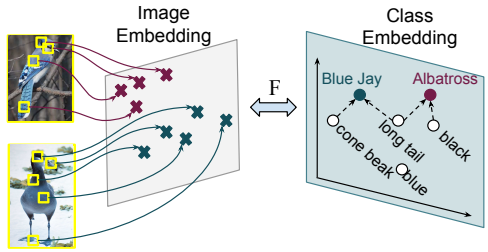
Xian et.al., CVPR'16 & CVPR'17



Linear compatibility function: large errors (left).
Piecewise-linear: significantly improves results (right).

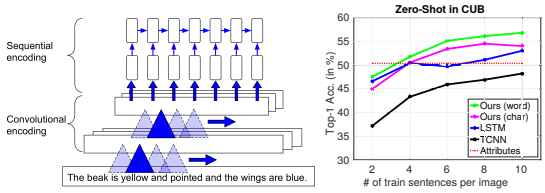
Multi-Cue Zero-Shot Learning with Strong Supervision

Akata et.al., CVPR'16



Attributes: costly but good, W2V: cheap but weak.
Strong visual supervision: to compensate weak W2V.

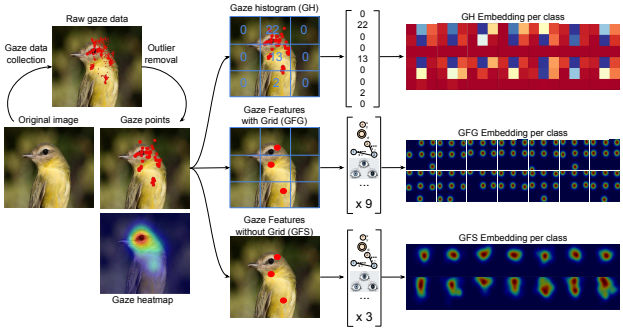
Learning Deep Representations of Fine-Grained Visual Descriptions Reed et.al., CVPR'16



CNN-RNN: fast + models sequence of words or characters
With >4 sentences: outperforms SoA with attributes

Gaze Embeddings for Zero-Shot Image Classification

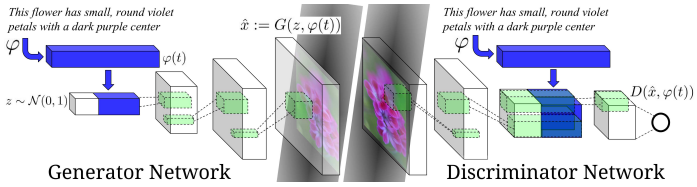
Karessli et.al., CVPR'17



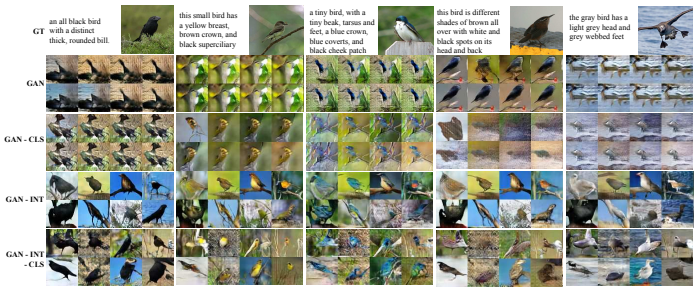
Generating: Vision + Language

Generarive Adversarial Text to Image Synthesis

Reed et.al. ICML'16

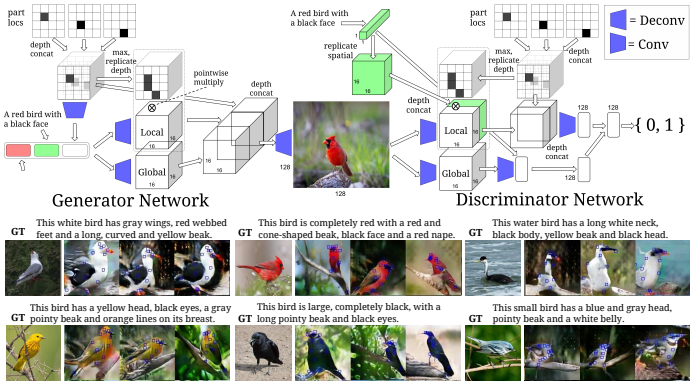


GAN conditioned on sentences: real/fake, matching/not



Generates pixels from characters: intuitive
Language compensates lack of large # training images

Learning What and Where to Draw Reed et.al. NIPS'16

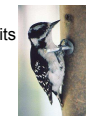


Generating Visual Explanations Hendricks et.al. ECCV'16

*This is a **Downy Woodpecker** because...*

D: this bird has a white breast black wings and a **red spot** on its head.
E: this is a black and white bird with a **red spot** on its crown.

*This is a **Downy Woodpecker** because...*



D: this bird has a white breast black wings and a **red spot** on its head.
E: this is a white bird with a black wing and a black and white striped head.

Class + image conditional LSTM & Reinforcement Loss
Learns to mention class-specific and visible properties

MAX-PLANCK-GESELLSCHAFT
max planck institut informatik