

Deep Generative Models

Part2: Imagination



Mohamed Elhoseiny
KAUST

mohamed.elhoseiny@kaust.edu.sa



Collaborators



Marcus



Othman



Antoine



Yannis



Manohar



Marc'aurelio



Camille



Yann



Scott



Adobe



Brian



Walter



Prithvijit



Ram



Stefan



Dhruv



Devi



Arslan



Sayna



Trevor



Tinne



Rahaf



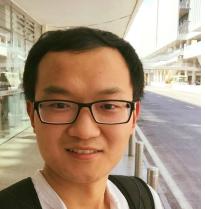
Francesca



Ji



Bingchen



Yitzhe



Ahmed



THE STATE UNIVERSITY
OF NEW JERSEY



People Imagine to Create



Mona Lisa(1503-1506)



Irises, Saint-Remy, Van Gogh (1889)

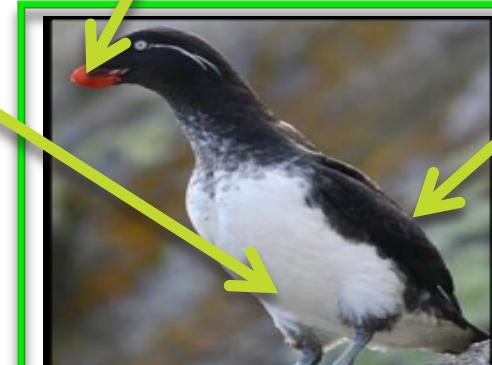


Starry Night, Van Gogh (1889)



Zero-Shot Learning from Text

Parakeet Auklet is a small bird that has a short **orange** peak. The bird's plumage is **dark** above and **white** below.

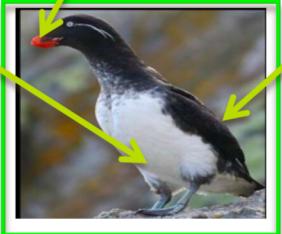




Imagination Inspired Vision

IMAGINE TO SEE

Parakeet Auklet is a small bird that has a short **orange** bill. The bird's plumage is **dark** above and **white** below.



IMAGINE TO CREATE





Machines Imagine to See, Why?

Parakeet Auklet is a small bird that has a short **orange** peak. The bird's plumage is **dark** above and **white** below.





Zero-Shot Visual Recognition

Generalize to Unseen Labels

Training (Seen) Classes:

- ▶ Black_footed_Albatross
- ▶ Crested_Auklet
- ▶ American_Crow

Test (Unseen Classes):

- ▶ Parakeet_Auklet
- ▶ Fish_Crow

“Side information”
“Representation of new classes ”

Recognizing Unseen Categories/ Zero-Shot Learning (ZSL)



Attribute Based Methods

[Lampert al., 2009, 2014]

[Farhadi, et al., 2009]

[Parikh, et al., 2010]

[Rohbrach et al., 2011])

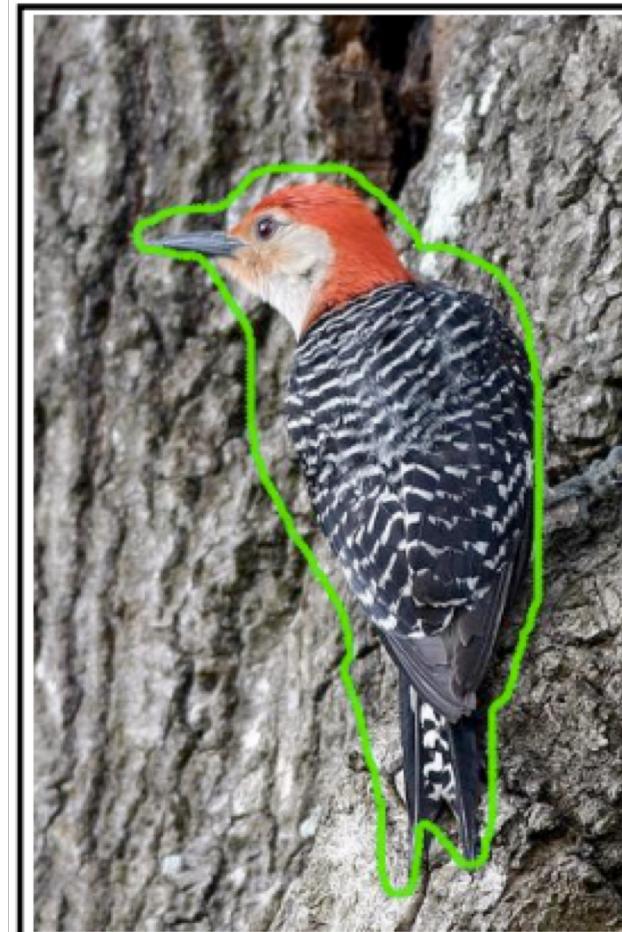
[Akata, et al., 2015]

[Xian, et al., 2017]

Drawbacks

- the dilemma of finding the best set of visual attributes
- Manual annotation for hundreds of attributes per class/ image

Example Attributes as Side Information

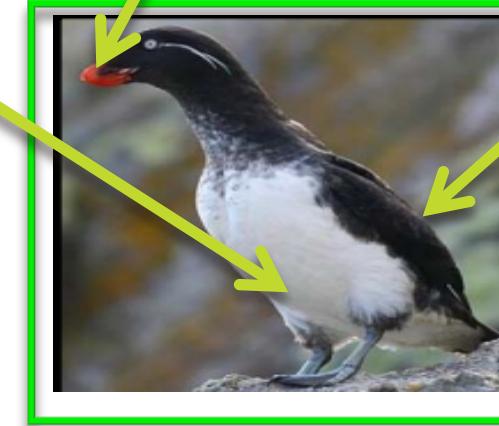


forehead_color	red	red	red
breast_pattern	multi-colored	solid	solid
breast_color	white	white/red	white
head_pattern	capped	capped	capped
back_color	white/black	white/black	white/black
wing_color	white/black	white/black	white/black
leg_color	buff	black	black
size	small	medium	medium
bill_shape	all-purpose	dagger	all-purpose
wing_shape	pointed	tapered	pointed
...
primary_color	black, red	white, black	white, black



Write a Classifier: ZSL from Pure Text Descriptions

Parakeet Auklet is a small bird that has a short **orange** peak. The bird's plumage is **dark** above and **white** below.



Interestingly, it is easy to collect such descriptions from sources like Wikipedia. However, it is a more challenging problem.

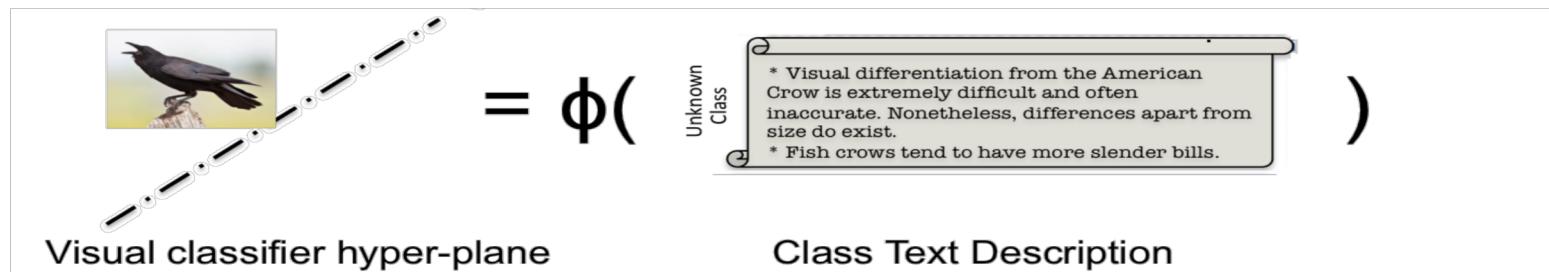


Linear Write a Classifier

- We assume a linear classifier $f_k(\mathbf{x}) = \mathbf{c}_k^T \cdot \mathbf{x}$

\mathbf{c}_k is a linear classifier for class k, \mathbf{x} is a visual feature vector appended by 1.

- The prediction in multiclass setting is $l^* = \arg \max_k f_k(\mathbf{x})$
- Could we explicitly predict classifier parameters of unseen classes from Unstructured Text?



Linear Write a Classifier: Learning W



t_1

* The Bobolink is a small New World blackbird and the only member of Dolichonyx.
* They often migrate in flocks, feeding on cultivated grains and rice, which leads to them being considered a pest by farmers in some areas.

t_N

* The Cardinals are a family of passerine birds found in North and South America. The South American cardinals in the genus Paroaria are placed in another family, the Thraupidae.

t_*

* Visual differentiation from the American Crow is extremely difficult and often inaccurate. Nonetheless, differences apart from size do exist.
* Fish crows tend to have more slender bills.

Known Object Classes

Unknown Classes

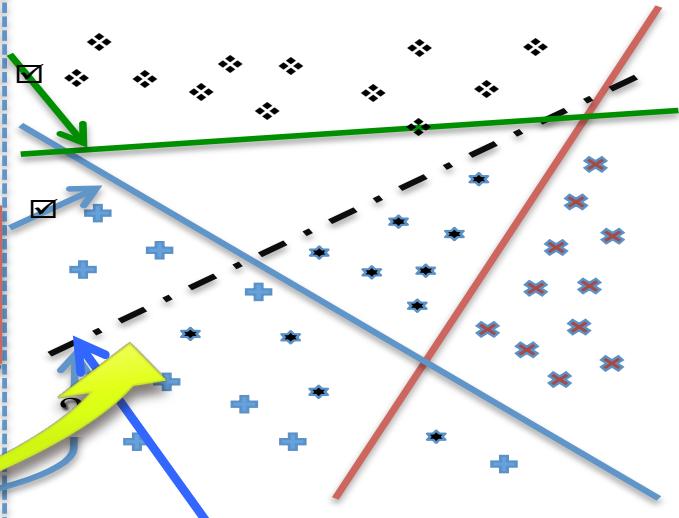


Fish Crow:

N/A

W

Visual Classifier Space



$$t_i^T W x_j > 1 \text{ if } t_i \text{ and } x_j \text{ belong to the same class, } t_i^T W x_j < -1 \text{ otherwise}$$

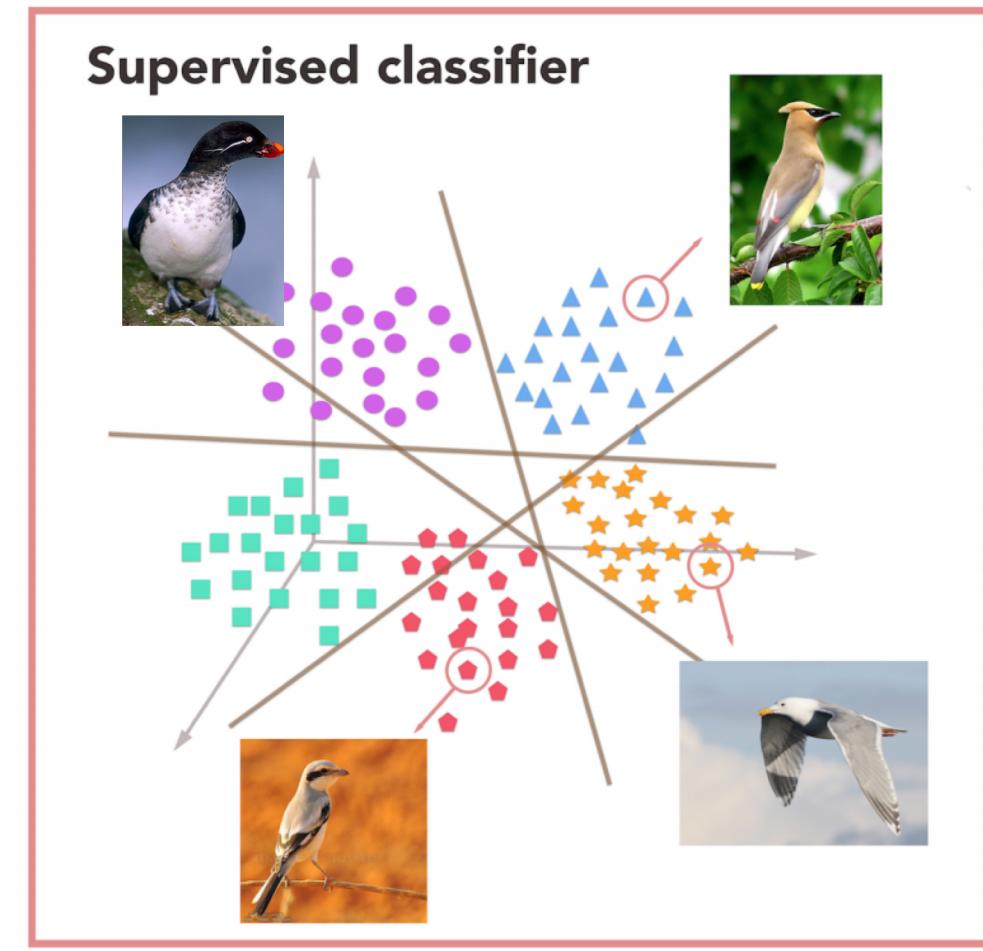
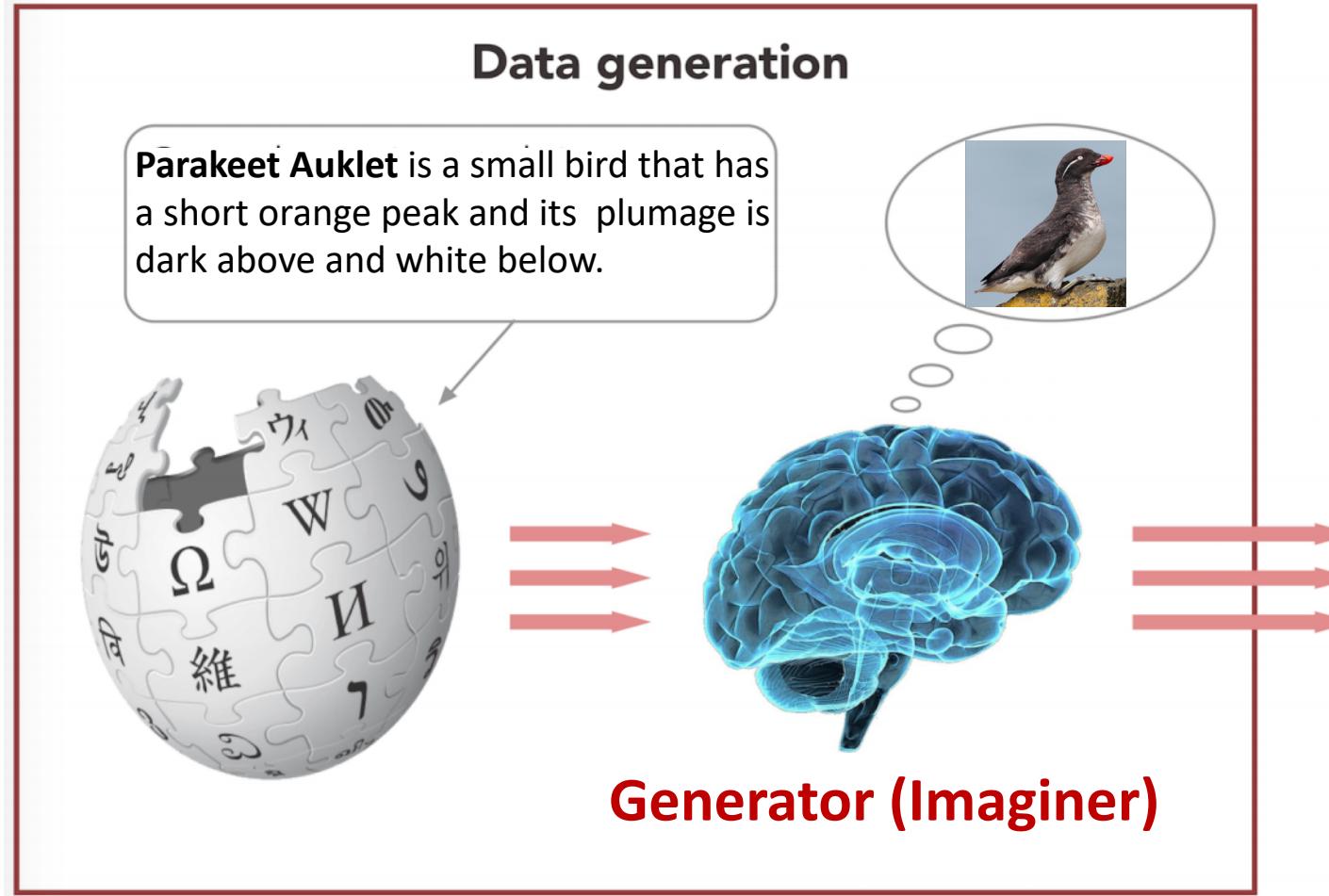


We haven't explicitly modeled imagination

Parakeet Auklet is a small bird that has an short **orange** peak. The bird's plumage is **dark** above and **white** below.



Imaginative Visual Classifier from Wikipedia Description



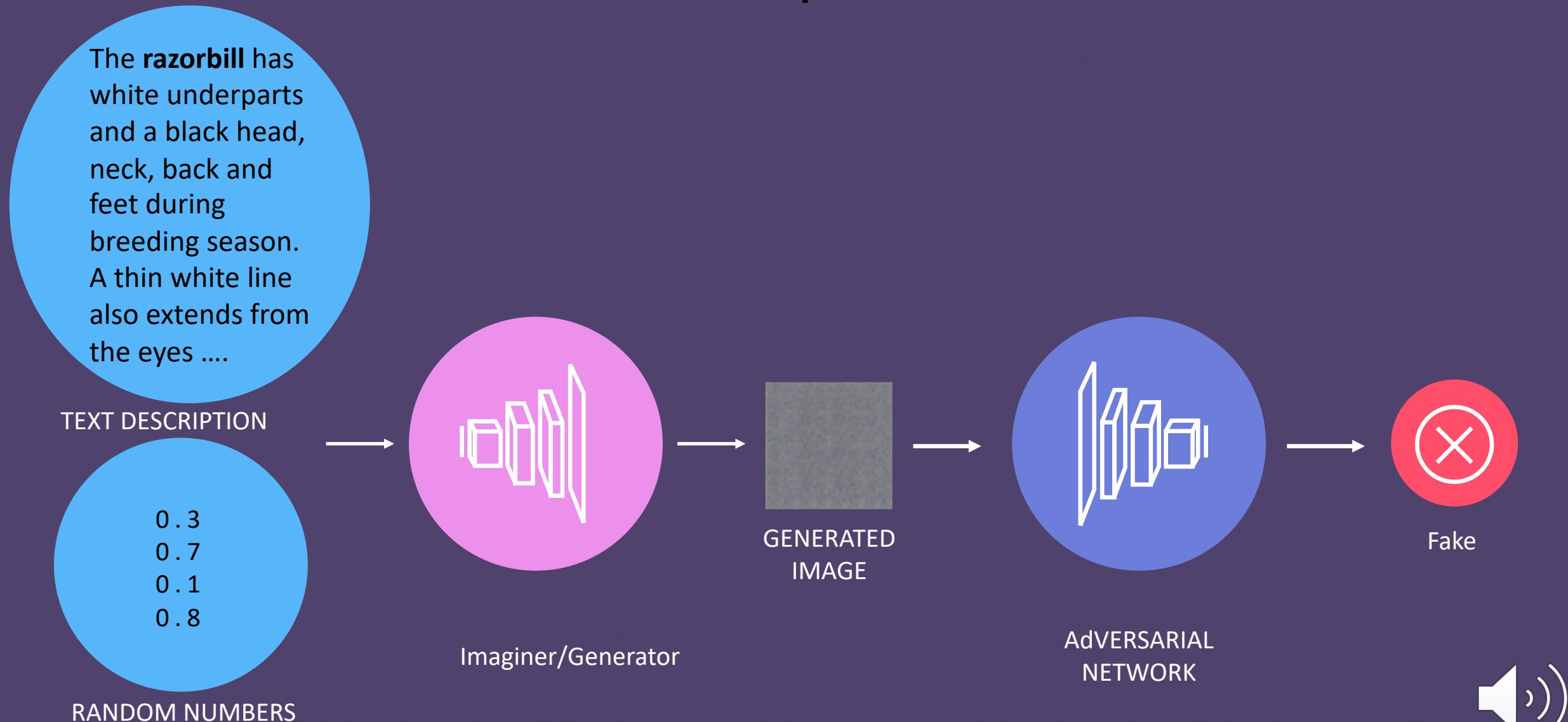


Imaginative Write a Classifier

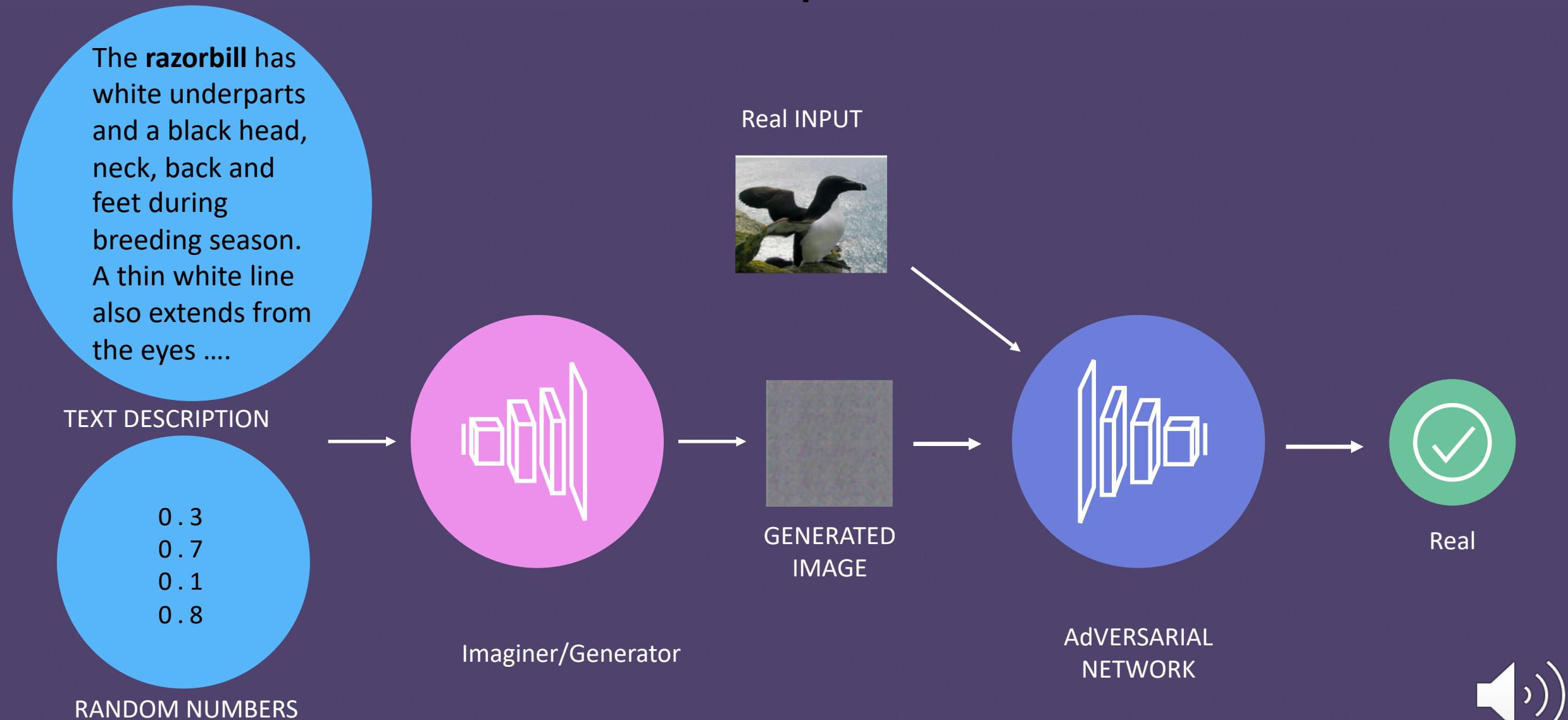
Generative Adversarial Networks (GAN)

[Godfellow et al., NIPS, 2014]

Imaginative Visual Classifier from Wikipedia Description



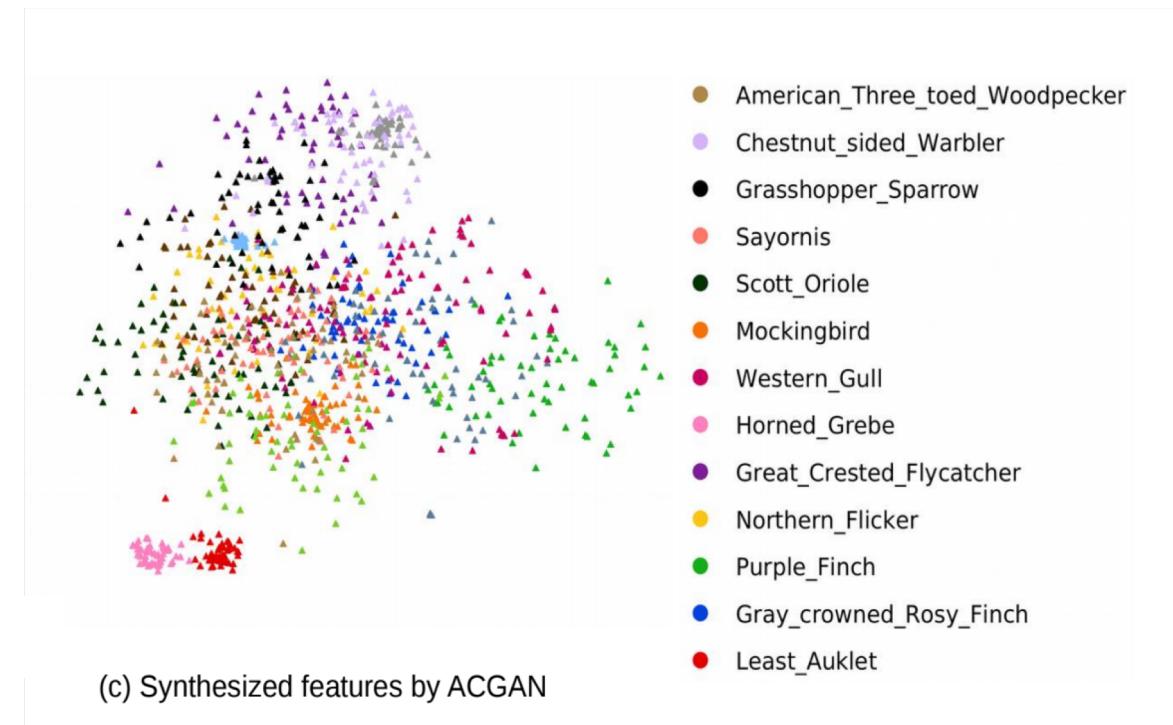
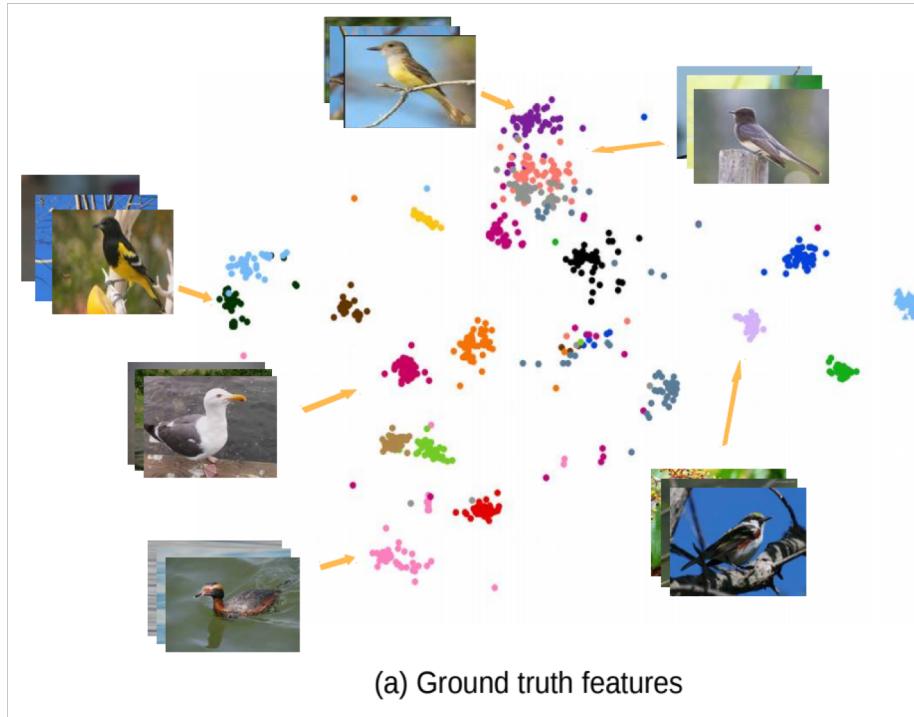
Imaginative Visual Classifier from Wikipedia Description





Imaginative Write a Classifier

Directly Applying Vanilla GAN does not work.

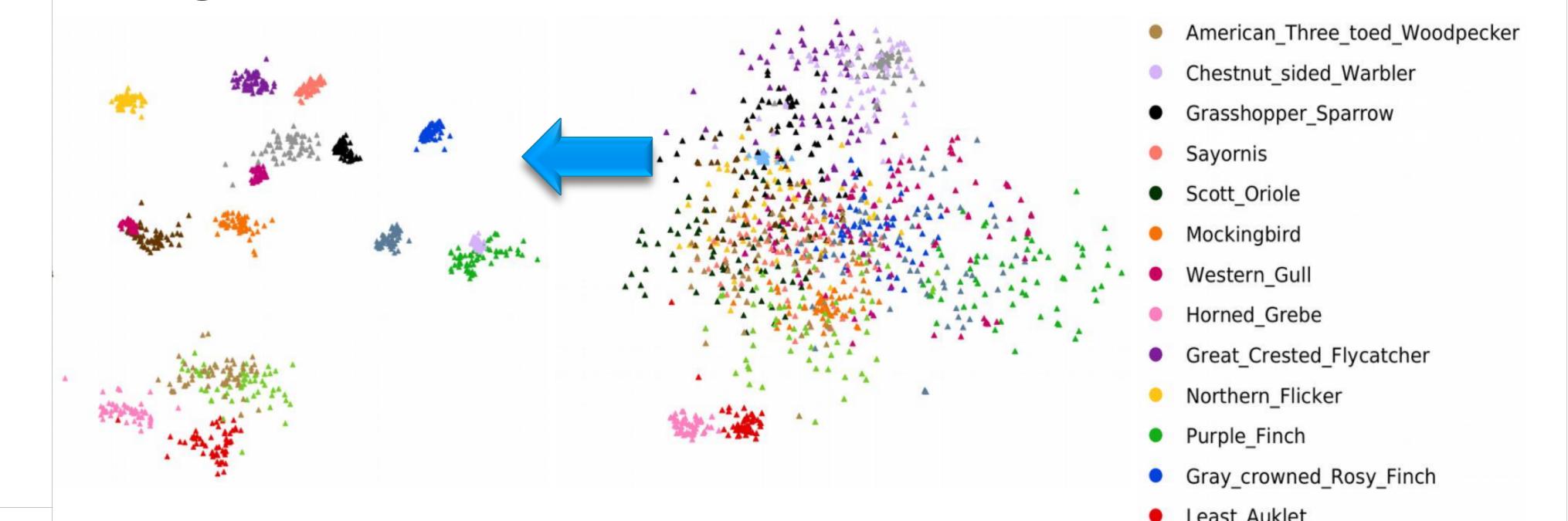


[Odena, et al, ICML, 2017]



Imaginative Write a Classifier

Visual Pivot Regularizer

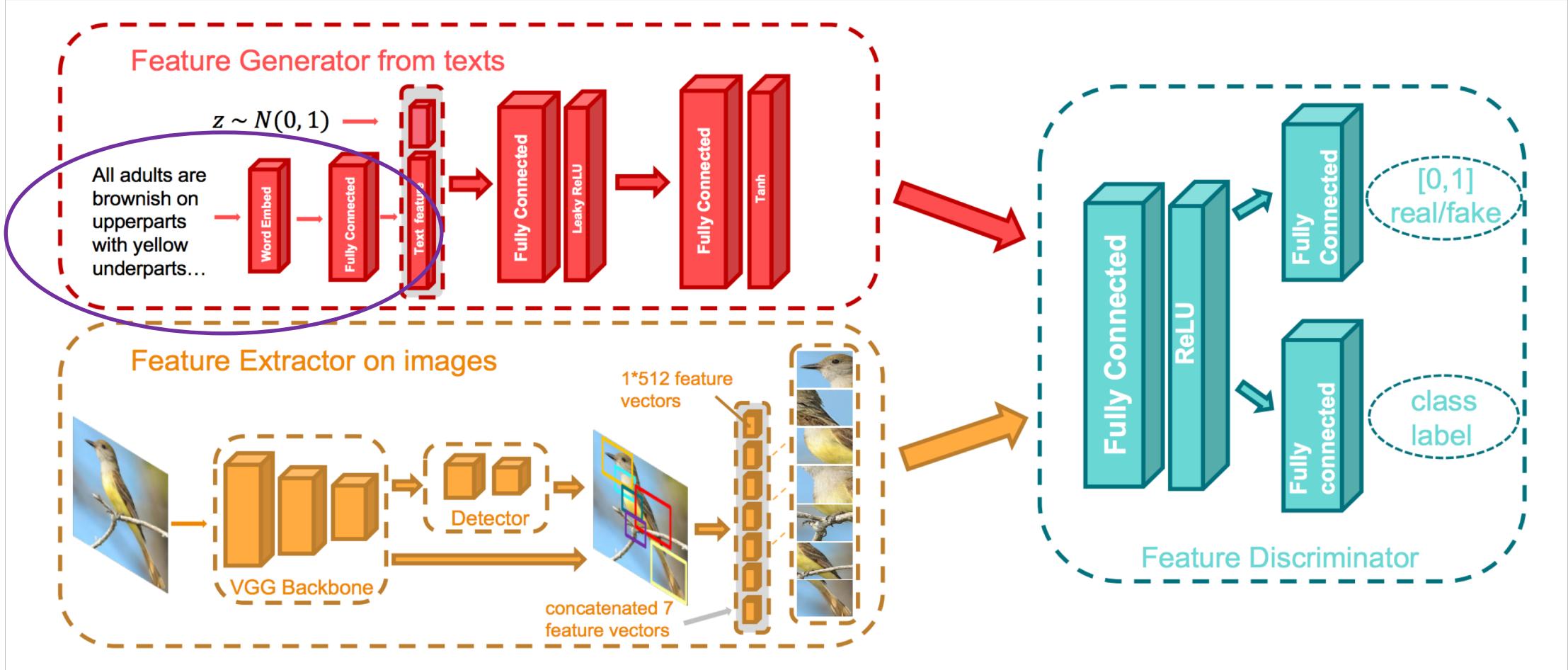


$$L_e = \frac{1}{C} \sum_{c=1}^C \left\| \mathbb{E}_{\tilde{x}_c \sim p_g^c} [\tilde{x}_c] - \mathbb{E}_{x_c \sim p_{data}^c} [x_c] \right\|^2,$$



Imaginative Write a Classifier

Note the Noise Suppression Layer on Top of Text





Wiki-CUB benchmark

- We extracted textual description available as augmentations of the CUB dataset of 200 species and 11000 images.
- 150 categories for training and 50 for testing.

* The Bobolink is a small New World blackbird and the only member of Dolichonyx.
* They often migrate in flocks, feeding on cultivated grains and rice, which leads to them being considered a pest by farmers in some areas.

Bobolink :



* The Cardinals are a family of passerine birds found in North and South America. The South American cardinals in the genus Paroaria are placed in another family, the Thraupidae.

Cardinals :

