Imagination Inspired Vision

KAUST
King Abdullah University of Science and Technology

Mohamed Elhoseiny
KAUST
mohamed.elhoseiny@kaust.edu.sa

Baidu Research
Silicon Valley AI Lab (SVAIL)

Collaborators

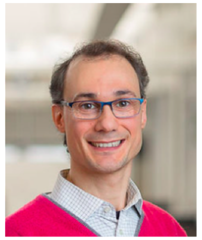

Marcus


Othman


Antoine


Yannis


Manohar


Marc'aurelio


Camille


Yann

facebook


Scott

Adobe

Georgia Tech

UNIVERSITY OF OXFORD

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH


Brian


Walter


Prithvijit


Ram


Stefan


Dhruv


Devi


Arslan


Sayna


Trevor

KU LEUVEN

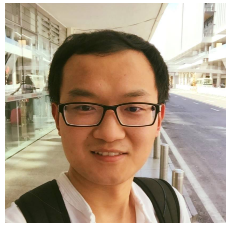

Tinne


Rahaf


Francesca


Ji

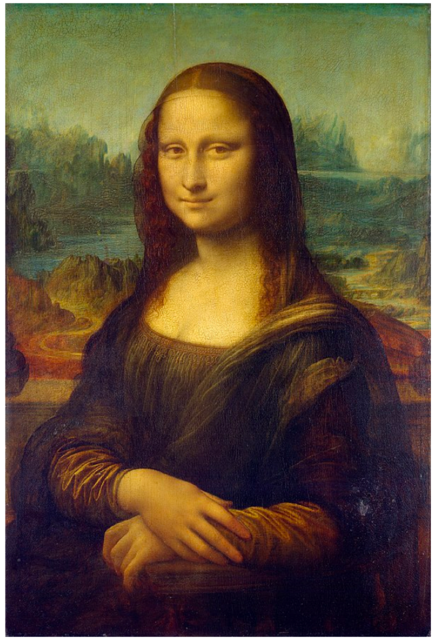

Bingchen


Yitzhe


Ahmed

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

People Imagine to Create



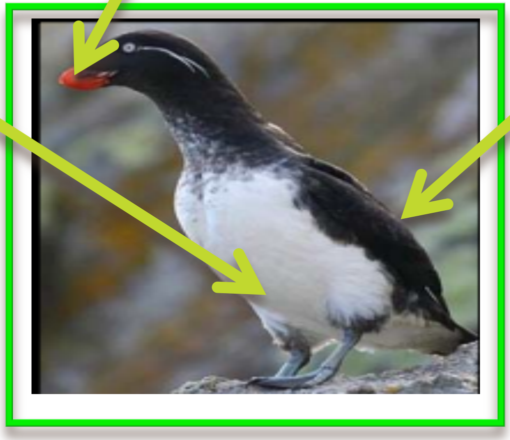
Mona Lisa(1503-1506)

Irises, Saint-Remy,Van Gogh (1889)

Starry Night, Van Gogh (1889)

Zero-Shot Learning from Text

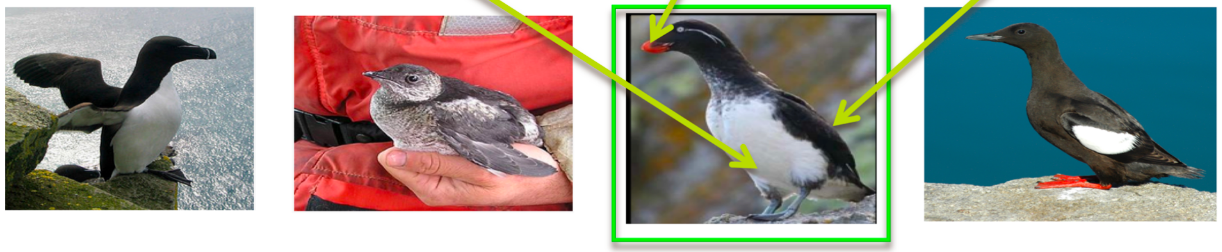
Parakeet Auklet is a small bird that has a short **orange** peak. The bird's plumage is **dark** above and **white** below.



Imagination Inspired Vision

IMAGINE TO SEE

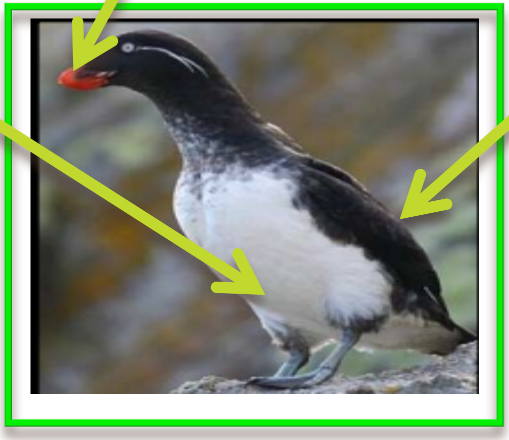
Parakeet Auklet is a small bird that has an short **orange** bill. The bird's plumage is **dark** above and **white** below.



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)

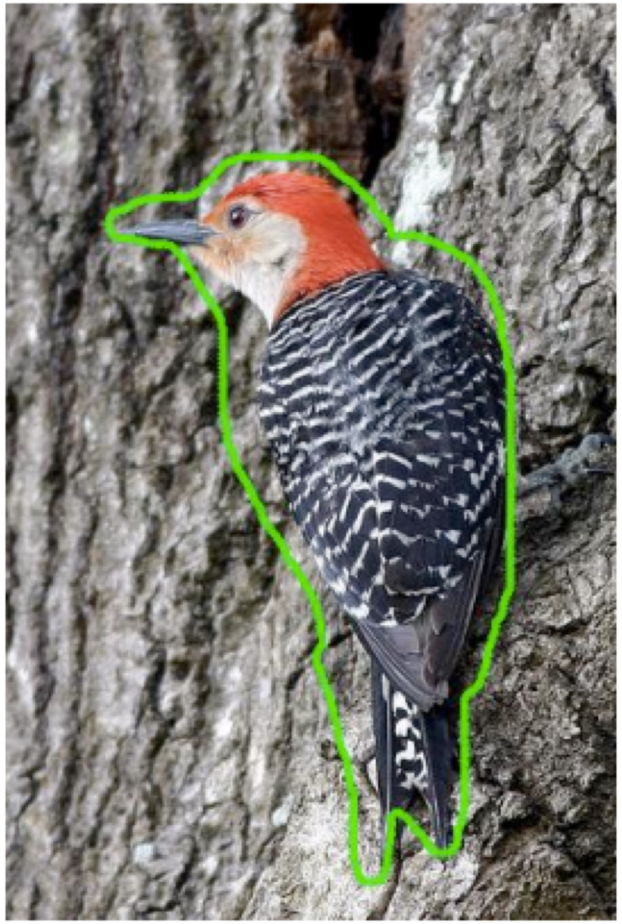
Attribute Based Methods

[Lampert al., 2009, 2014]
[Farhadi, et al., 2009]
[Parikh, et al., 2010]
[Rohbrach et al., 2011)]
[Akata, etal, 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

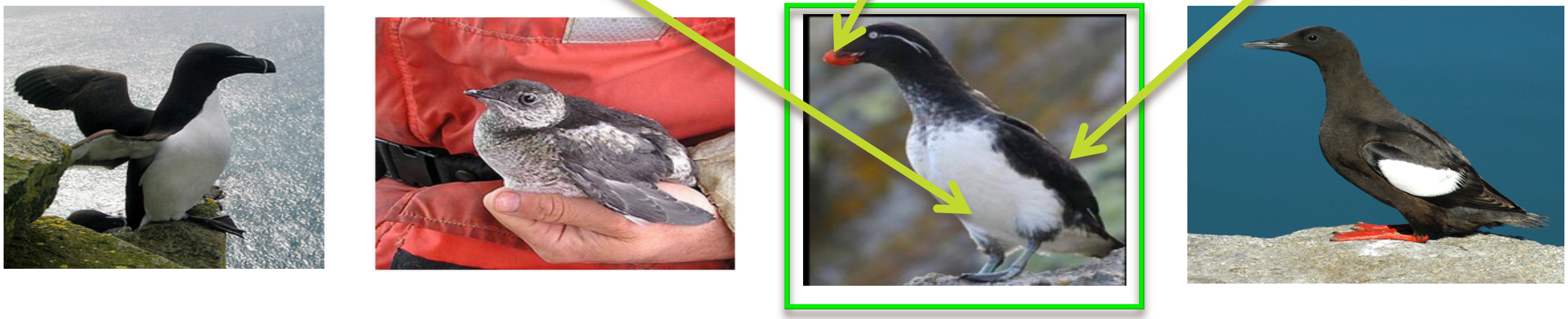


forehead_color	**red**	red	**red**
breast_pattern	multi-colored	solid	solid
breast_color	white	white/red	white
head_pattern	**capped**	capped	**capped**
back_color	**white/ black**	**white/ black**	**white/ black**
wing_color	**white/ black**	white/black	white/black
leg_color	buff	black	black
size	**small**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

Welinder et al. 2010, Birds Dataset

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.

[Elhoseiny et al., ICCV, 2013]

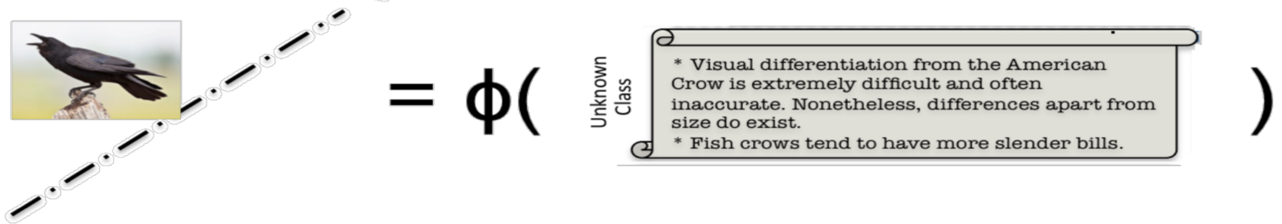
Linear Write a Classifier

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

  $\mathbf{c}_k$ 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?**

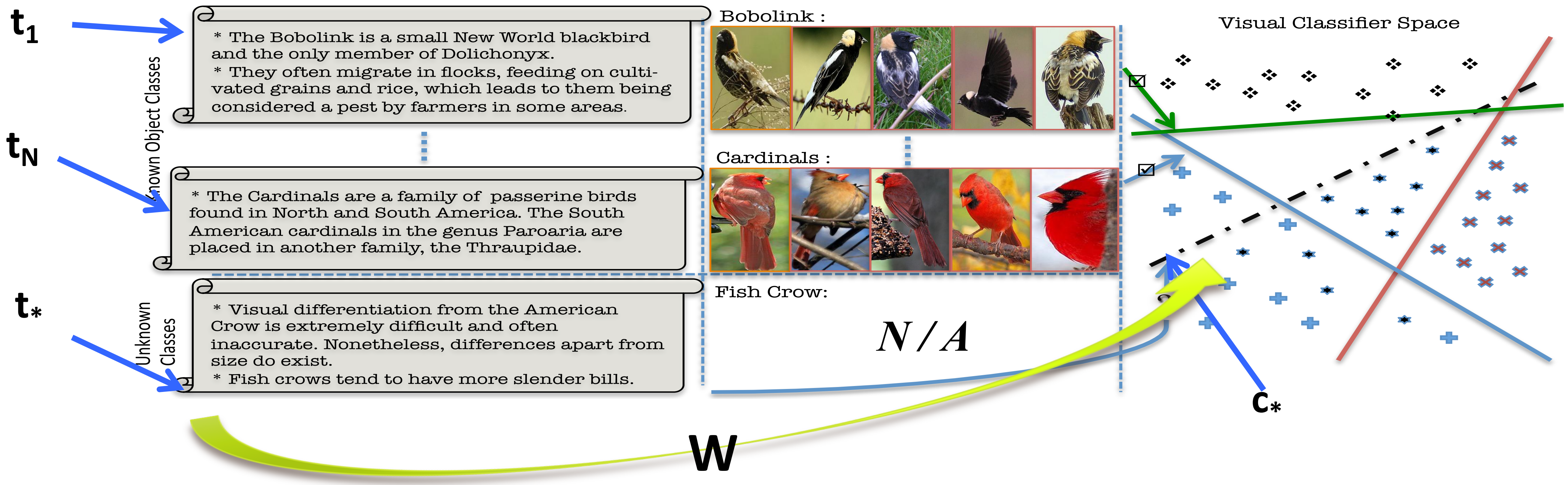


Visual classifier hyper-plane

Class Text Description

[Elhoseiny et al., ICCV, 2013]

Linear Write a Classifier: Learning W

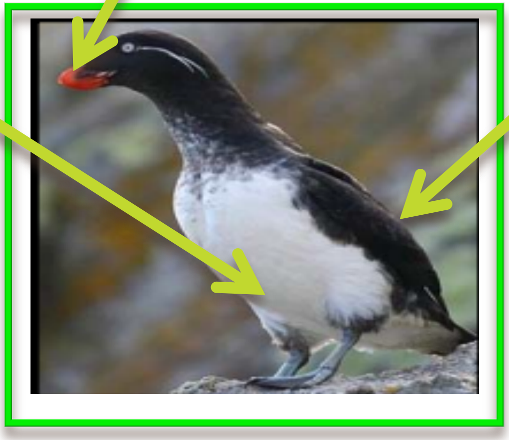


$\mathbf{t}_i^T \mathbf{W}\, \mathbf{x}_j > 1$ if $\mathbf{t}_i$ and $\mathbf{x}_j$ belong to the same class, $\mathbf{t}_i^T \mathbf{W}\, \mathbf{x}_j < -1$ otherwise

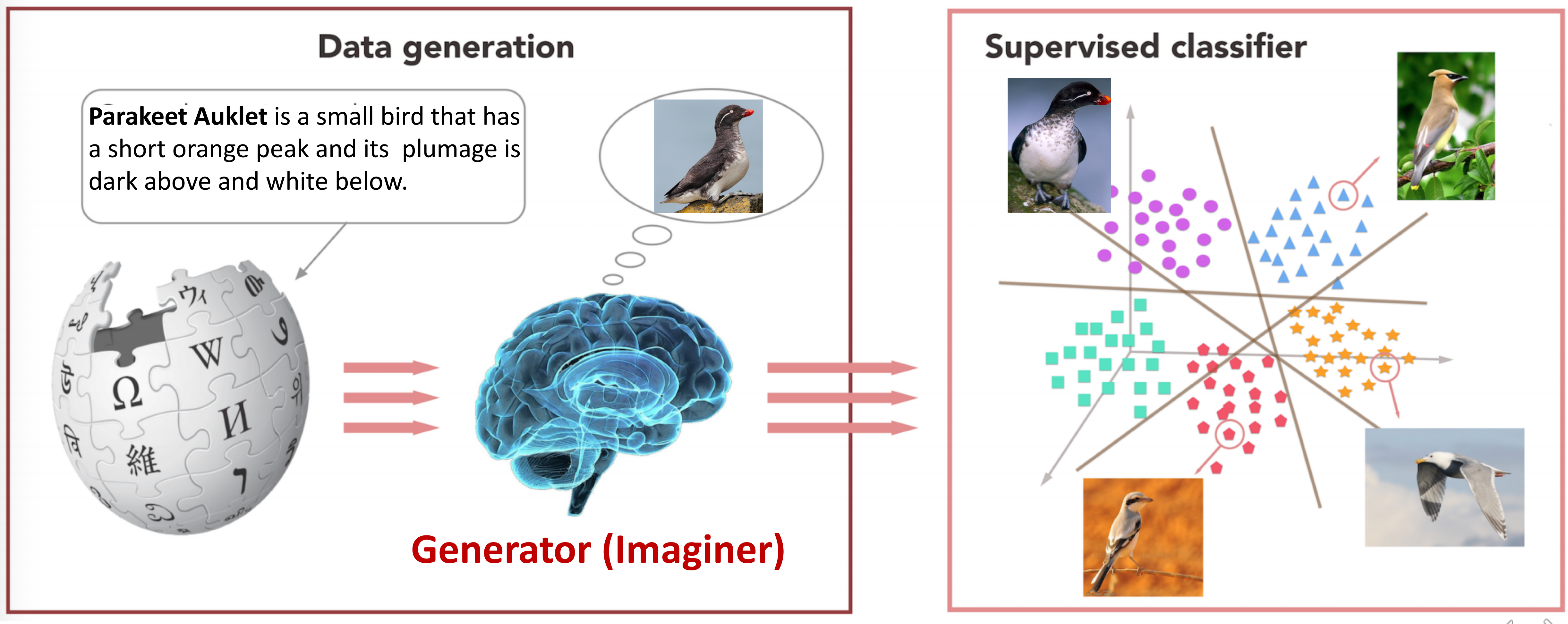
[Elhoseiny et al., ICCV, 2013]

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

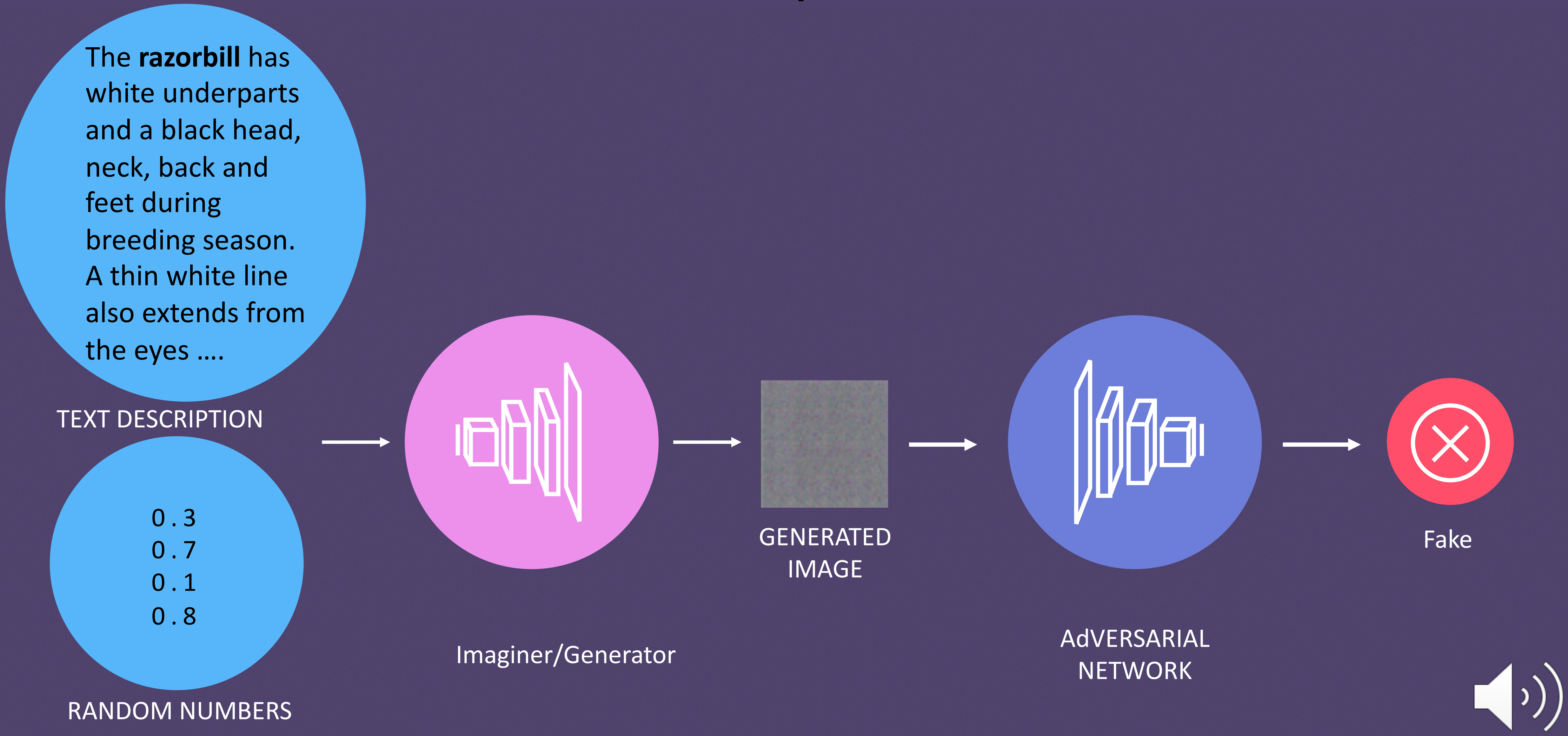


[ZELE et al., CVPR, 2018]

Generative Adversarial Networks (GAN)

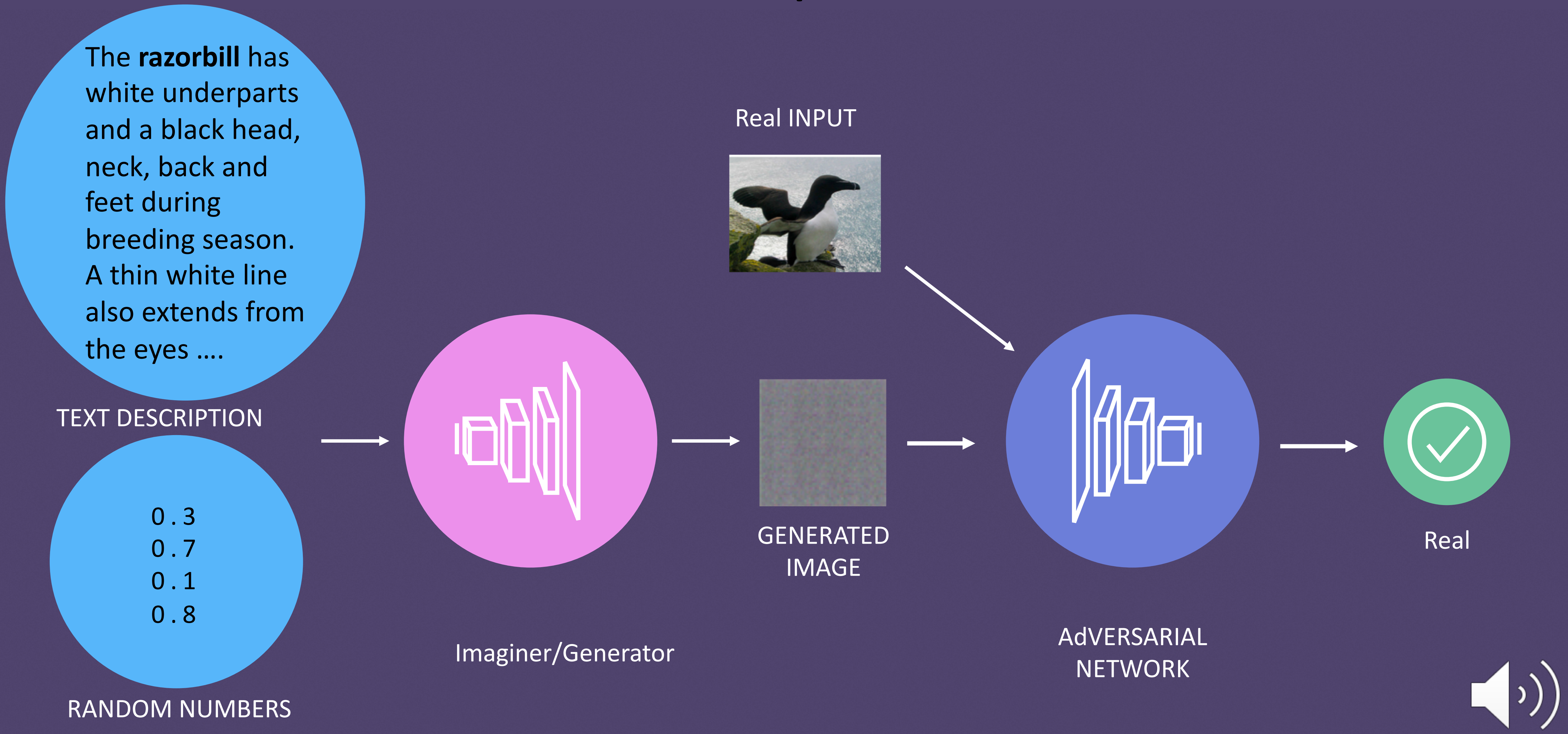
[Godfellow et al., NIPS, 2014]

Imaginative Visual Classifier from Wikipedia Description



[Godfellow et al., NIPS, 2014; Odena, et al, ICML, 2017 ; ZELE et al., CVPR, 2018]

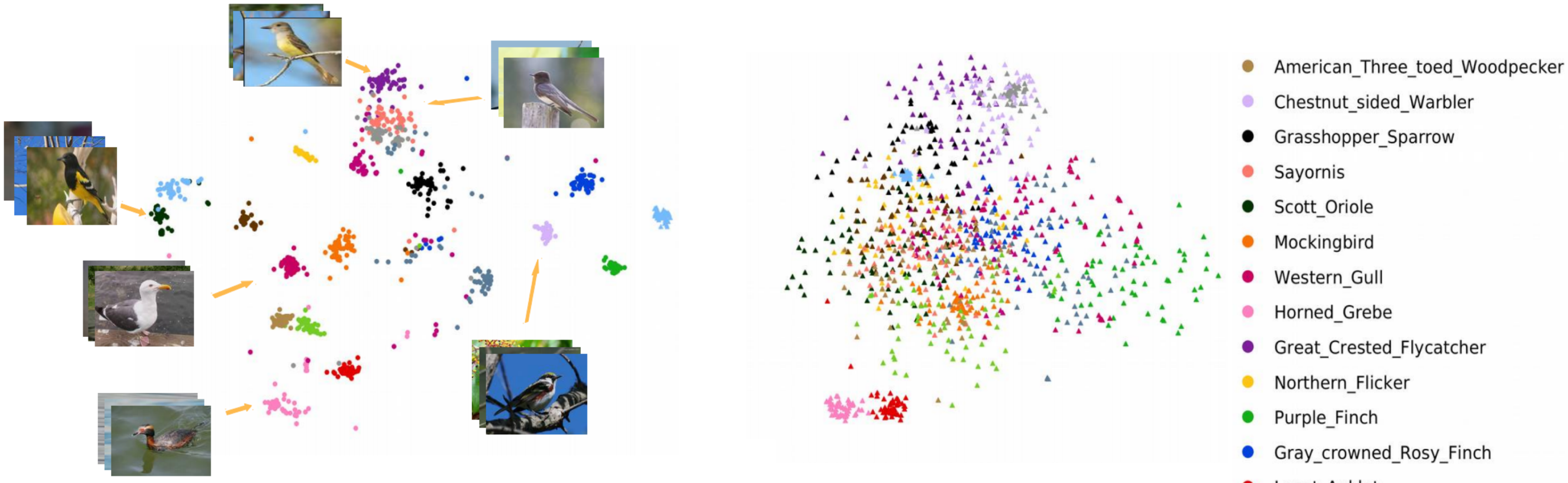
Imaginative Visual Classifier from Wikipedia Description



[Godfellow et al., NIPS, 2014; Odena, et al, ICML, 2017 ; ZELE et al., CVPR, 2018]

Imaginative Write a Classifier

Directly Applying Vanilla GAN does not work.



(a) Ground truth features

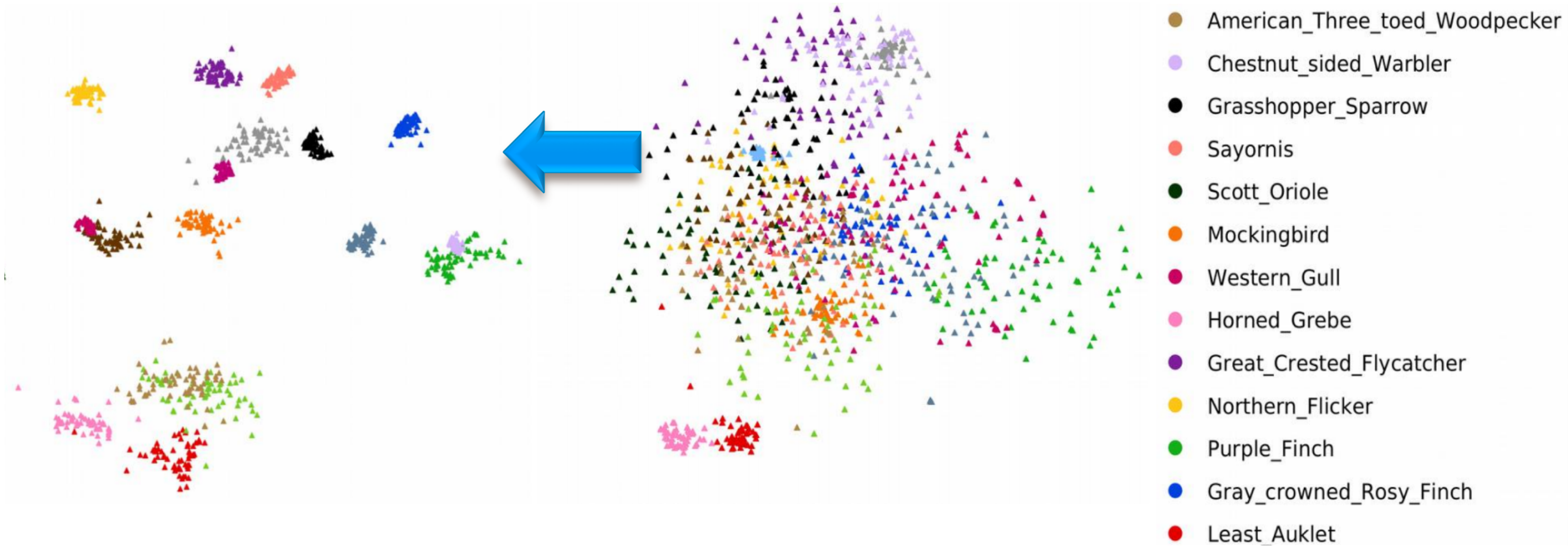
(c) Synthesized features by ACGAN

[Odena, et al, ICML, 2017]

[ZELE et al., CVPR, 2018]

Imaginative Write a Classifier

Visual Pivot Regularizer

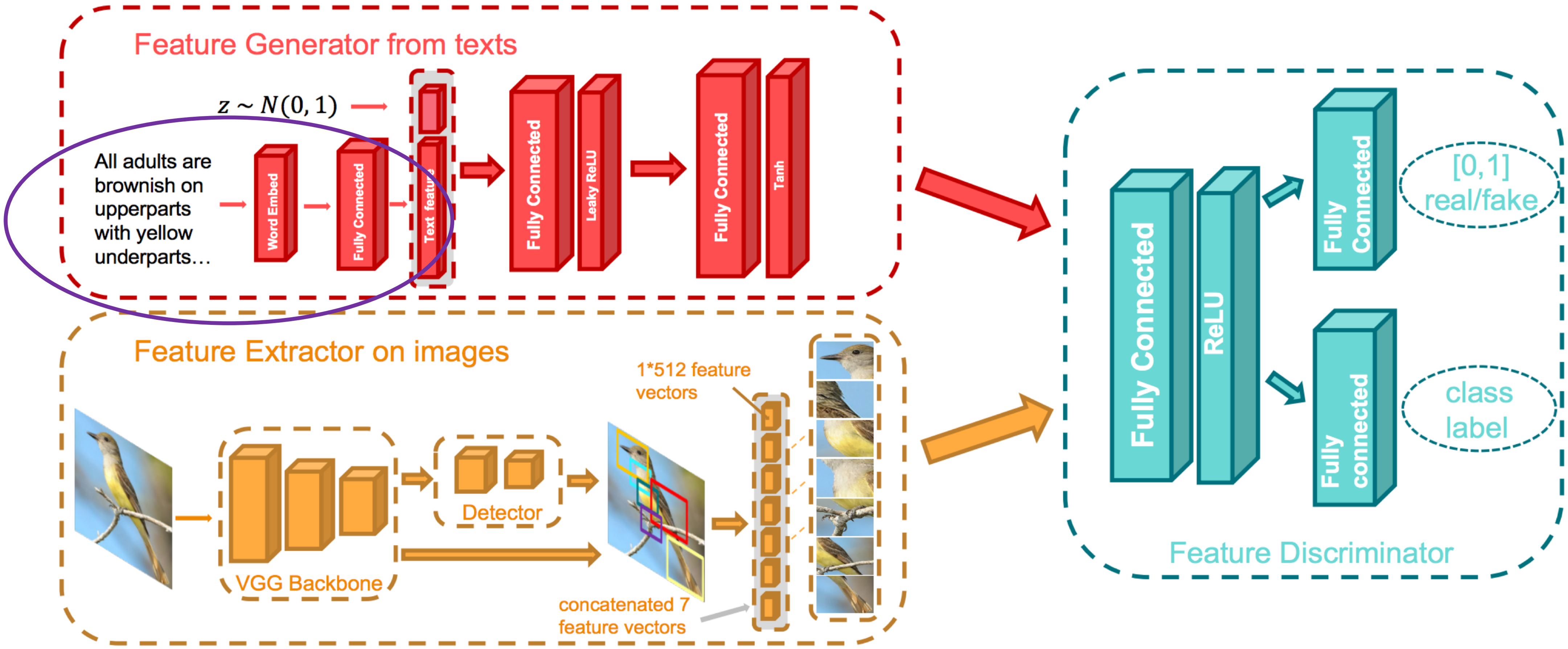


$$L_e = \frac{1}{C} \sum_{c=1}^{C} ||\mathbb{E}_{\tilde{x}_c \sim p_g^c}[\tilde{x}_c] - \mathbb{E}_{x_c \sim p_{data}^c}[x_c]||^2,$$

[ZELE et al., CVPR, 2018]

Imaginative Write a Classifier

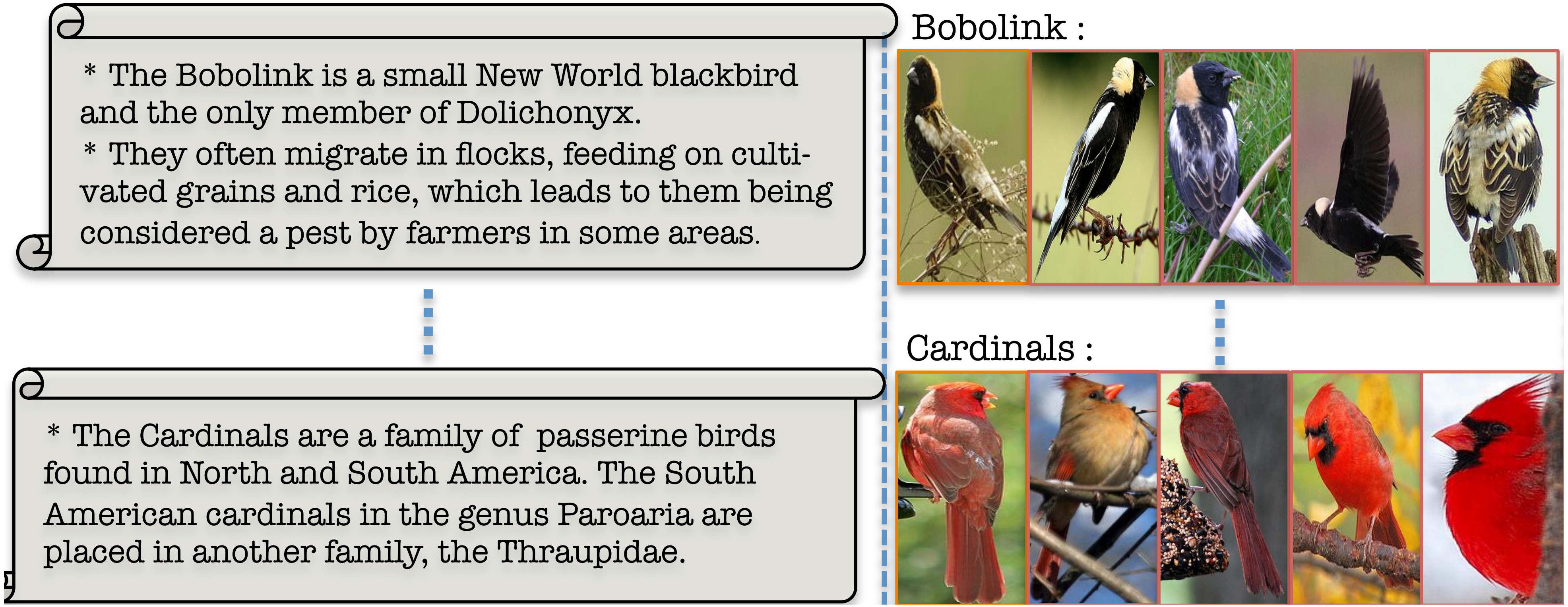
Note the Noise Suppression Layer on Top of Text



[ZELE et al., CVPR, 2018]

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.



Ablation Study

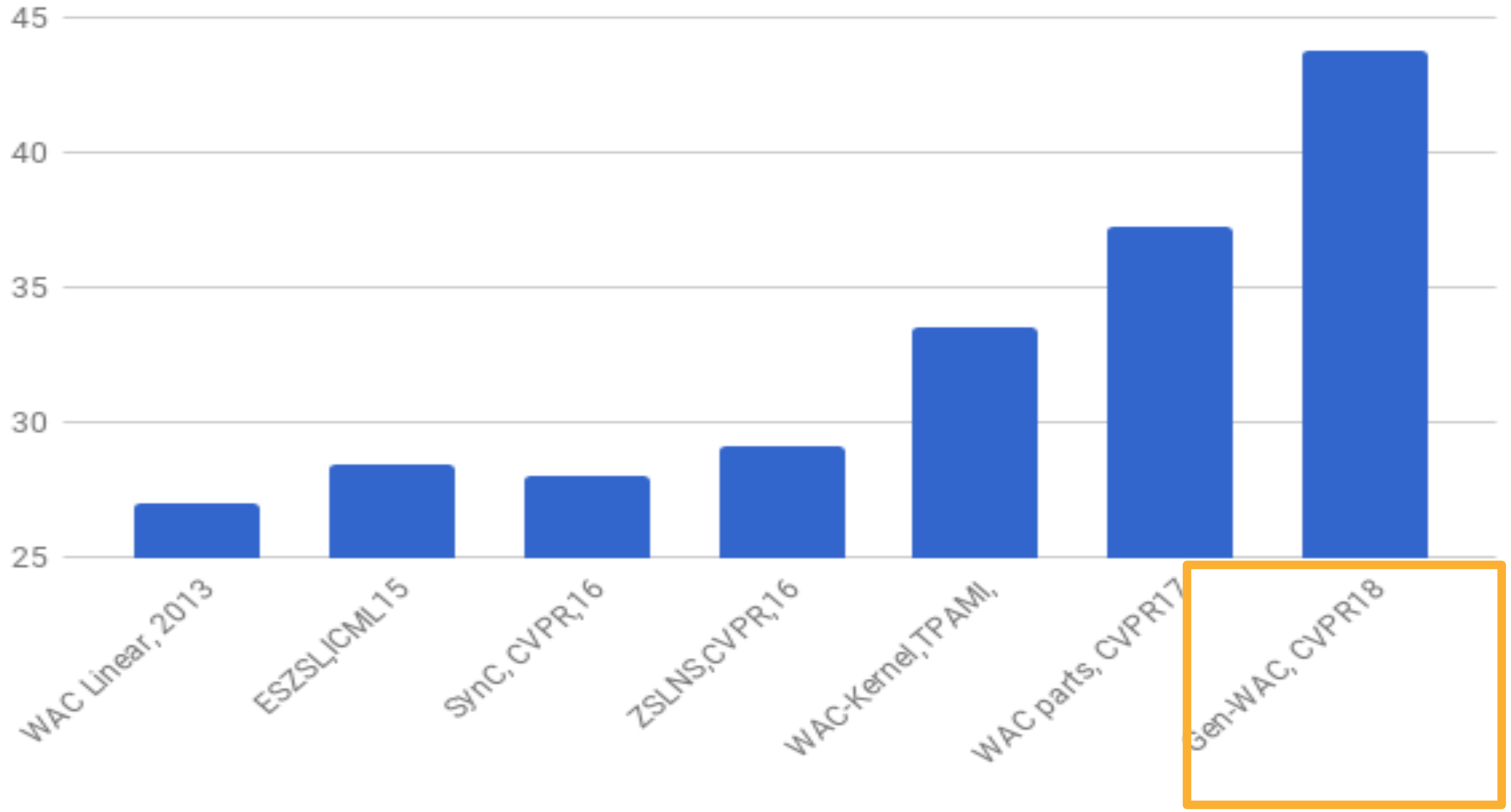
w/ FC means with Noise Suppression Layer .

	CUB		NAB	
methods	w/ FC	w/o FC	w/ FC	w/o FC
GAN-only	22.83	21.83	24.22	24.80
VP-only	28.52	26.76	25.75	23.42
Ours	**43.74**	40.85	**35.58**	32.94

[ZELE et al., CVPR, 2018]

Imaginative Write a Classifier

Comparison to the ZSL State of the Art



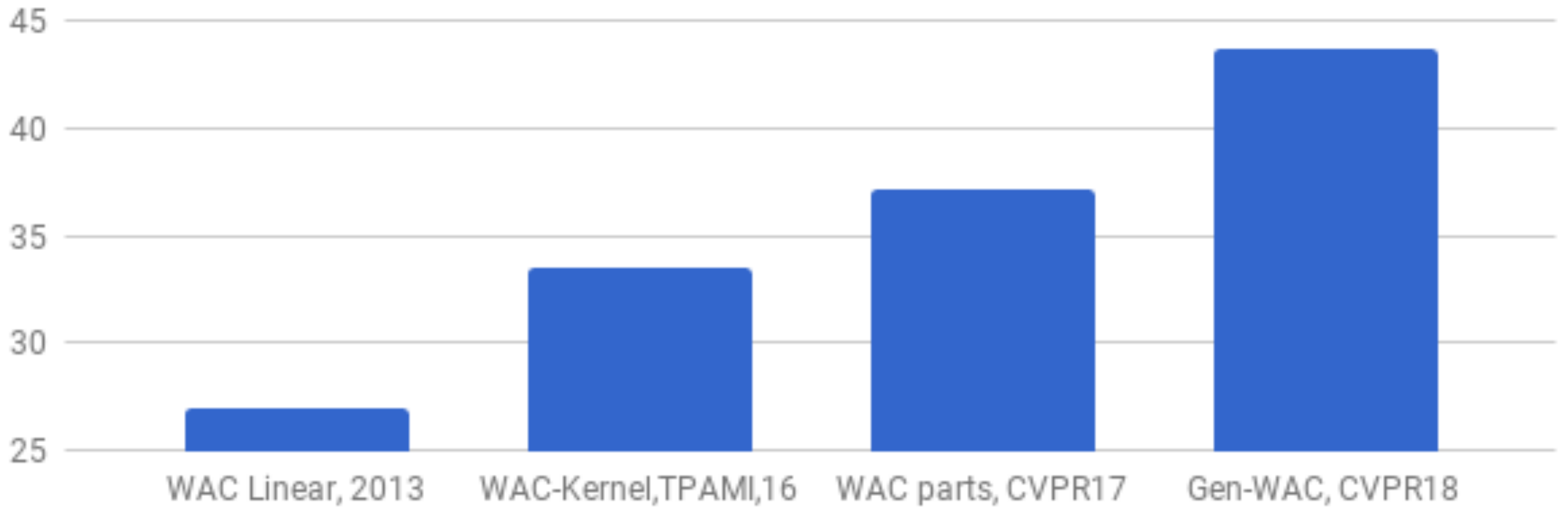
Parallel work in related problems

[Xian, etal, 2018]
[Bharath, etal, 2017]
[Wang etal, 2018]
[YZ etal, 2018]

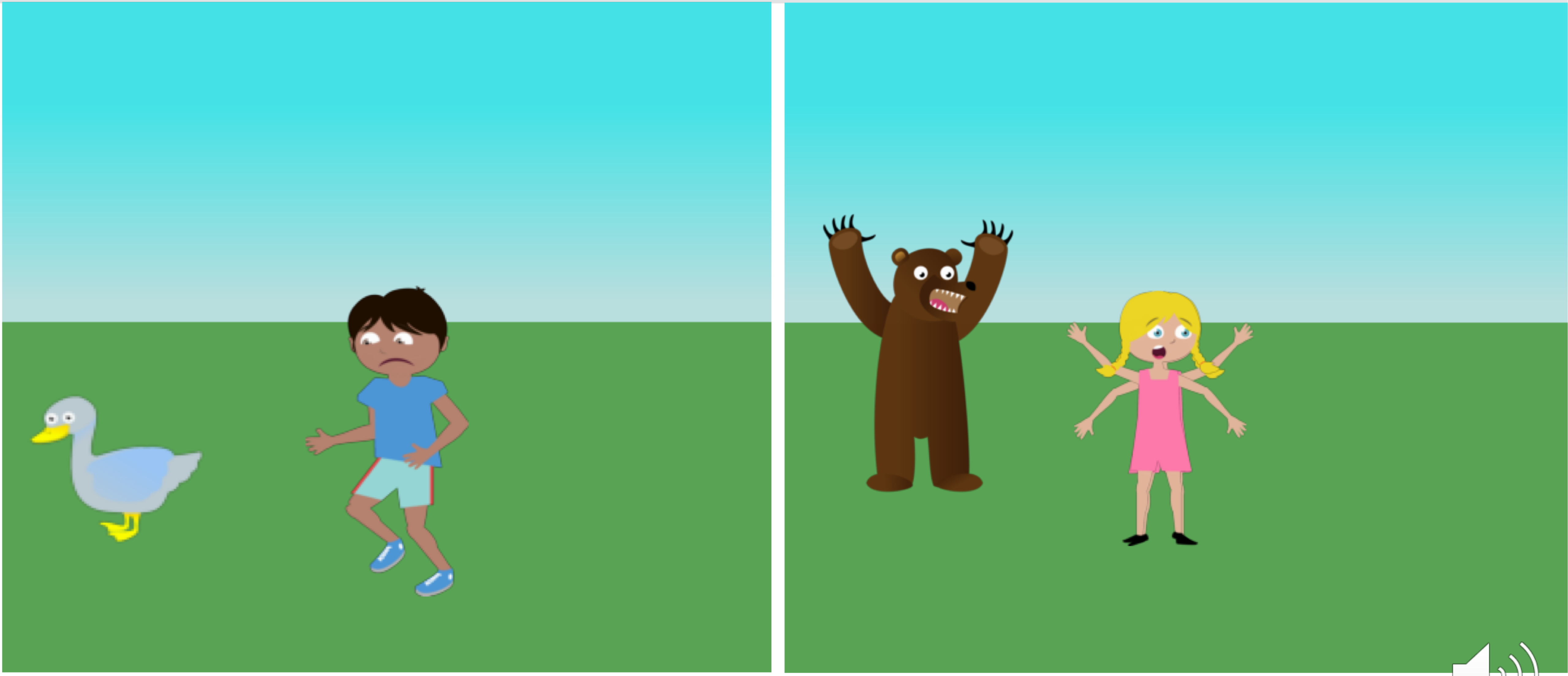
[ZELE et al., CVPR, 2018]

Write a Classifier (5 year Summary)

- [ESE, ICCV, 2013] : Linear approach with 26% on CUB
- [EES, TPAMI, 2016] Non non-linear kernel classifiers with 33.5% on CUB
- [EZE, CVPR, 2017] Modeling the parts notion with 37.2% on CUB
- [ZELE,CVPR, 2018] Modeling a visual imaginer from text helps with 43.7% on CUB



We often encounter this

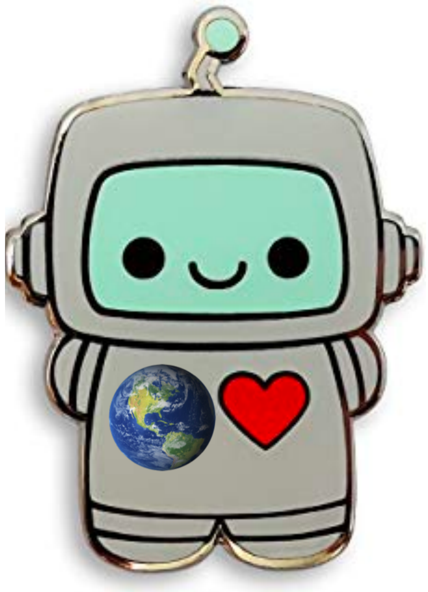


To visualize this scenario, I used cartoon characters from Abstract Scenes dataset (C. L. Zitnick and D. Parikh, 2013)

AI : An Additional Arm to help Mother Nature at UN

Adam

I saw a bird with an orange beak, what is it?



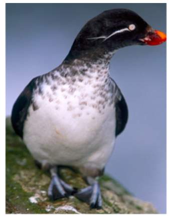
I am not quite sure, Is the bird's plumage dark above?

yes

Is it also white above?

yes

I think you are talking about "parakeet Auklet.", Is this it?

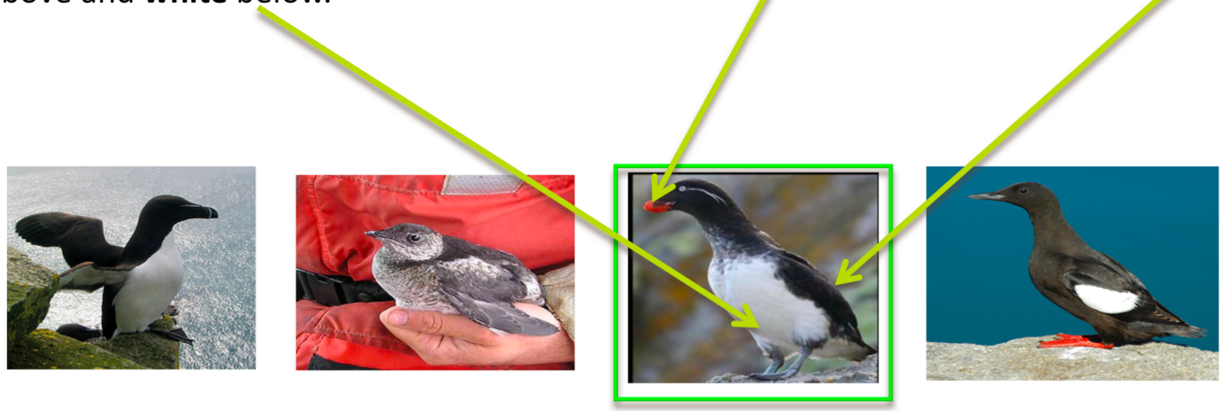


yes

Imagination Inspired Vision

IMAGINE TO SEE

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

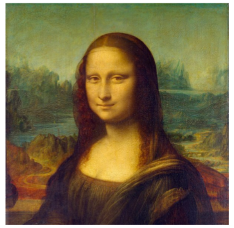


IMAGINE TO CREATE

Creative Adversarial Networks (CAN)

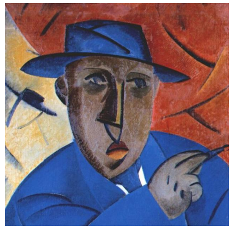
Creation from Random Numbers

High Renaissance

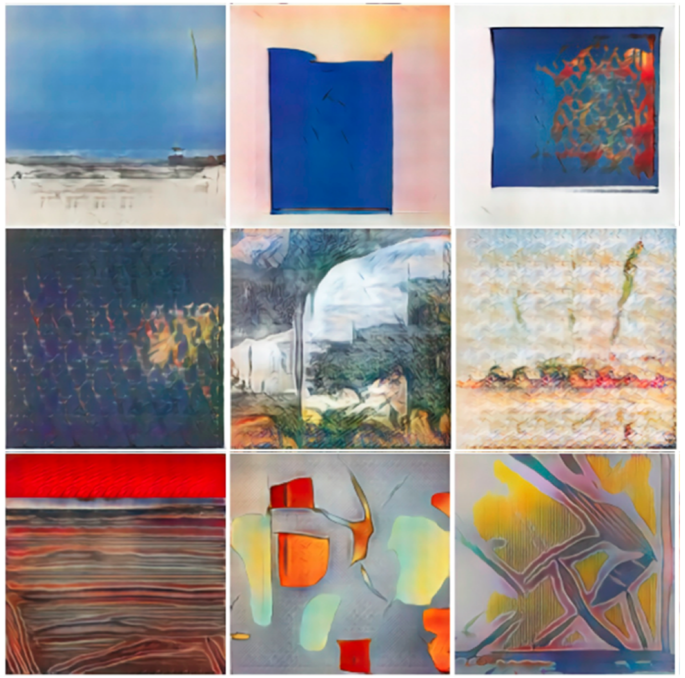


Abstract Art

Cubism

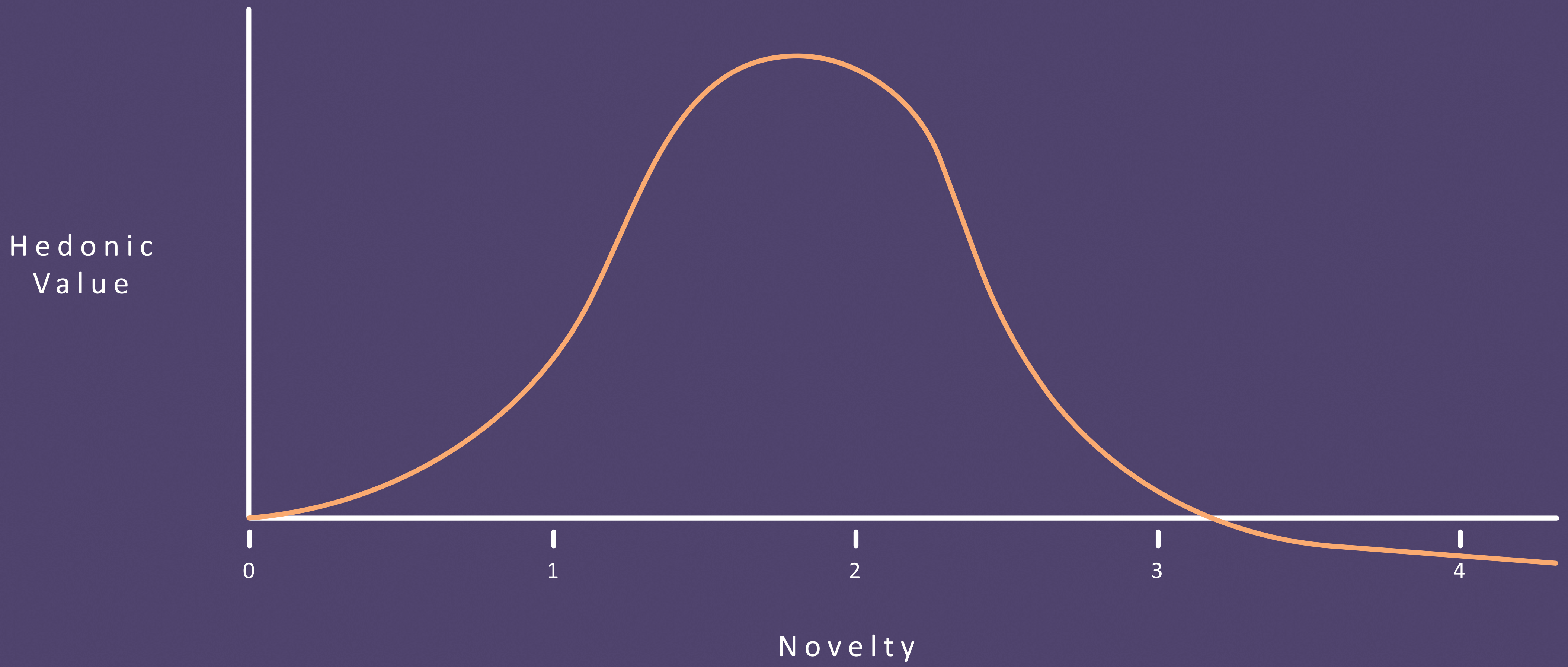


New style



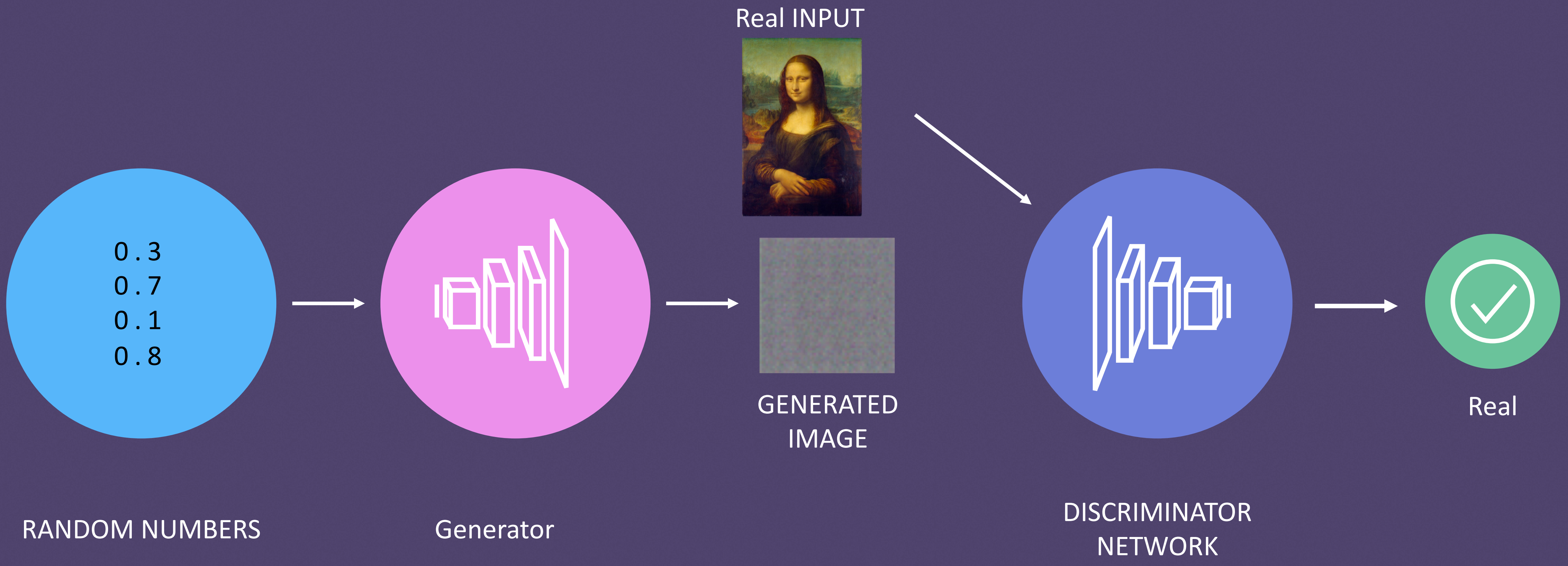
[ELEB., ICCC, 2017]

Principle of least effort: Wundt curve



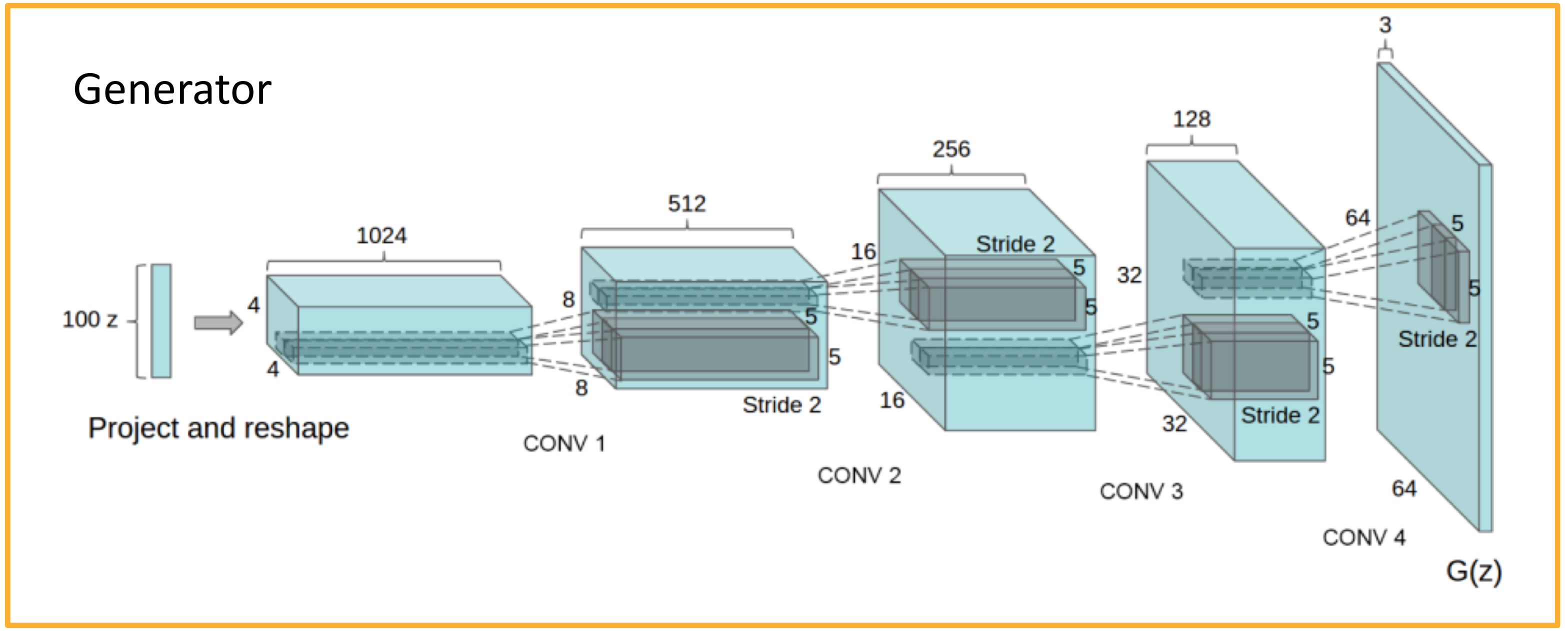
[Slide credit, our F8 presentation]

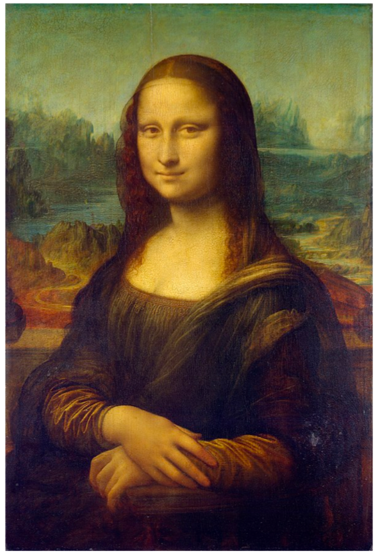
Generative Adversarial Networks



[Slide credit, our F8 presentation]

GAN has no motivation to be creative



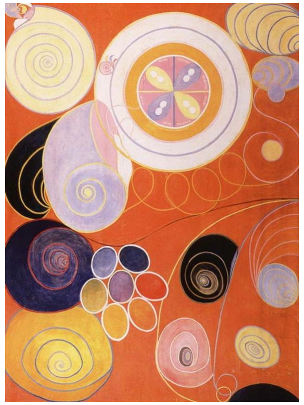


AI Creative Artist?

No, not creative.

Wiki Art 20 Style Classes and Modeling the deviation

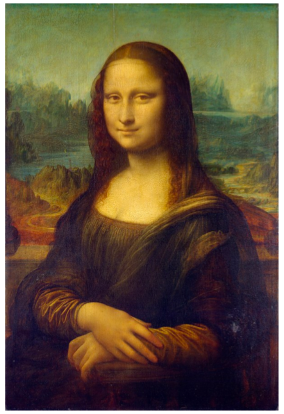
Abstract Art



Cubism

Impressionism

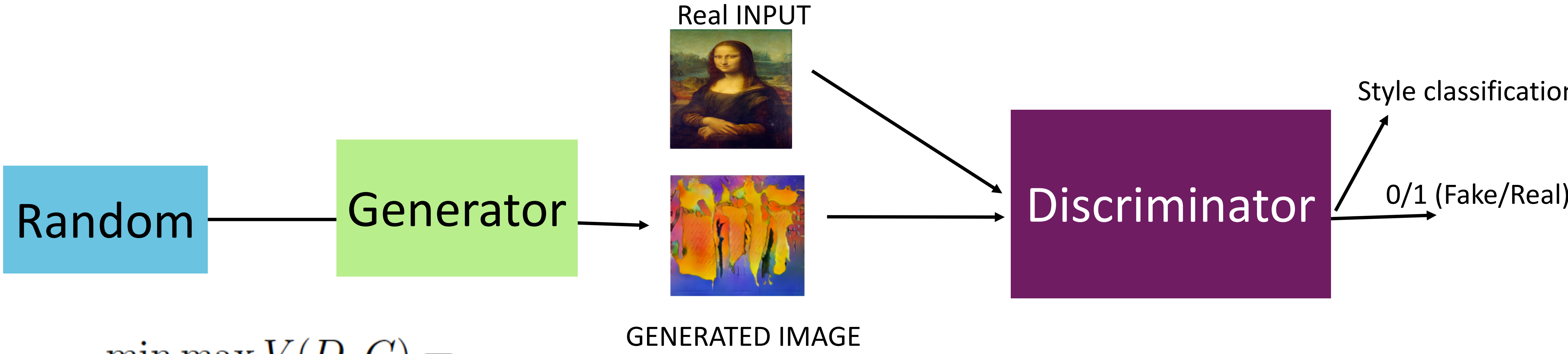
High Renaissance



……

[ELEB., ICCC, 2017]

Creative Adversarial Networks



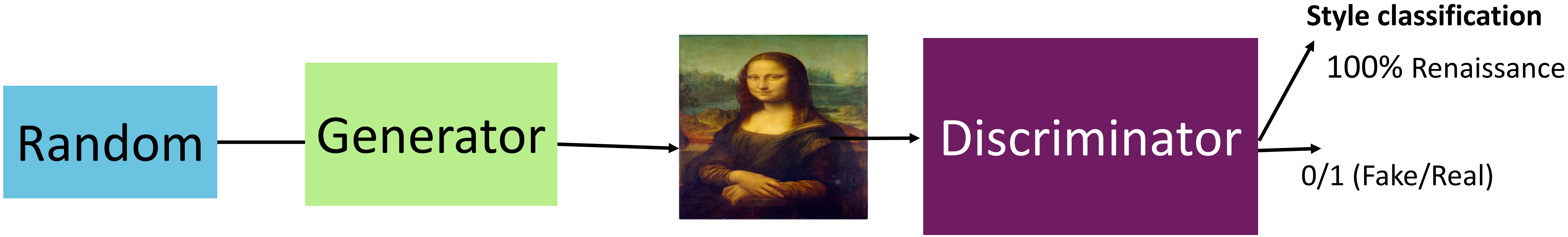
$$\min_{G} \max_{D} V(D,G) =$$
$$\mathbb{E}_{x,\hat{c}\sim p_{data}}[\log D_r(x) + \log D_c(c=\hat{c}|x)] +$$
$$\mathbb{E}_{z\sim p_z}[\log(1-D_r(G(z))) - \sum_{k=1}^{K}(\frac{1}{K}log(D_c(c_k|G(z))+$$
$$(1-\frac{1}{K})log(1-D_c(c_k|G(z)))],$$

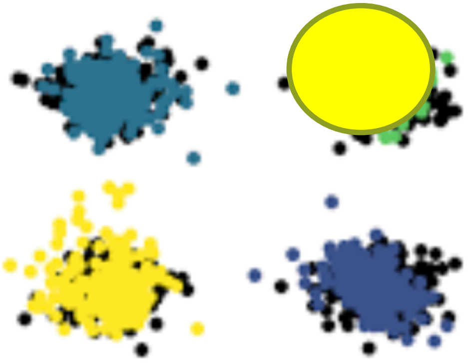
Creativity Loss

[ELEB., ICCC, 2017]

Creative Adversarial Networks

Low Style Ambiguity (low Entropy)= Low Creativity



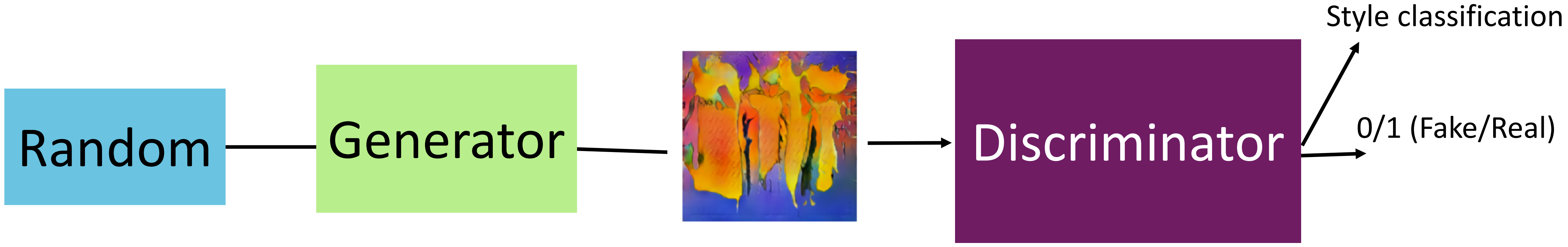


$$-\sum_{k=1}^{K}(\frac{1}{K}log(D_c(c_k|G(z))+$$
$$(1-\frac{1}{K})log(1-D_c(c_k|G(z)))],$$

[ELEB., ICCC, 2017]

Creative Adversarial Networks

High Style Ambiguity (high Entropy)= high Creativity



$$-\sum_{k=1}^{K}\Big(\frac{1}{K}log(D_c(c_k|G(z))+ (1-\frac{1}{K})log(1-D_c(c_k|G(z)))\Big],$$

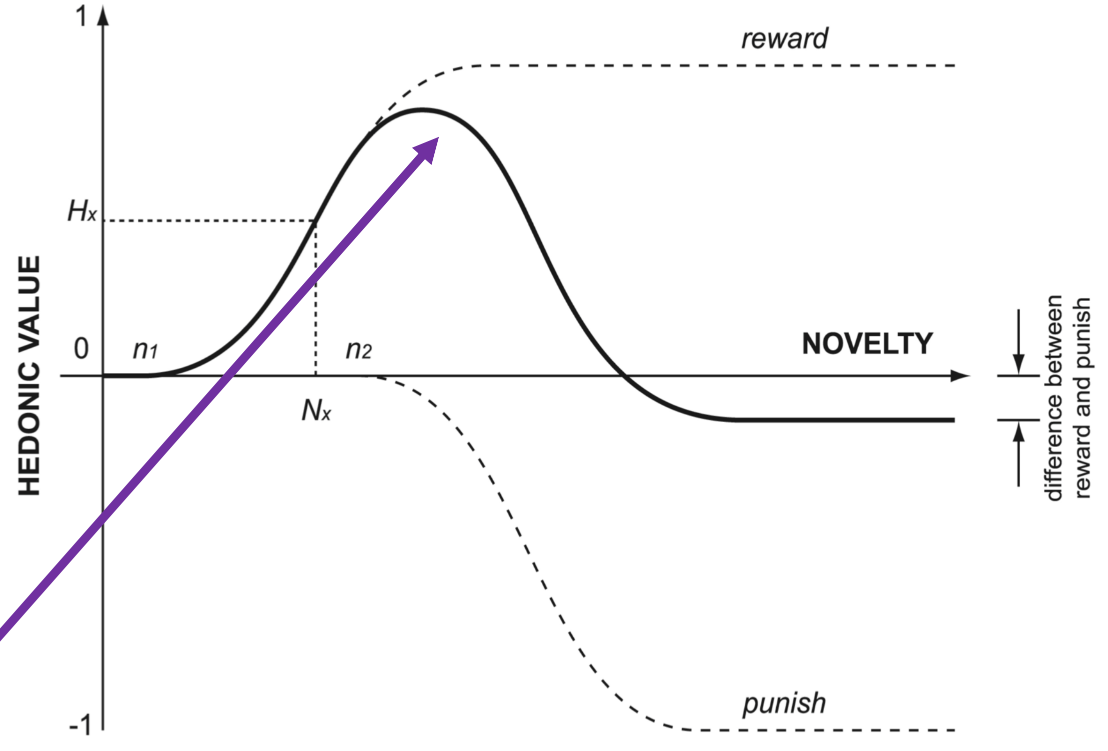
[ELEB., ICCC, 2017]

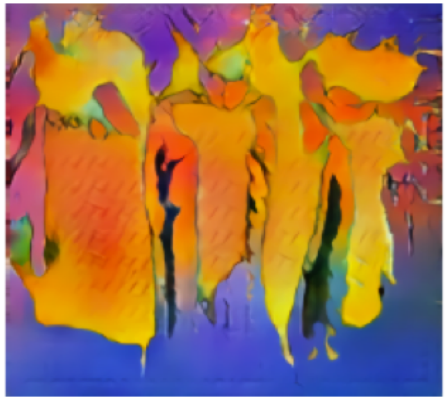
Creative Adversarial Networks

Loss and Connection to the Principle Of Least Effort

Colin Martindale (1943–2008)

$$\min_G \max_D V(D,G) = \mathbb{E}_{x,\hat{c}\sim p_{data}}[\log D_r(x) + \log D_c(c=\hat{c}|x)] + \mathbb{E}_{z\sim p_z}[\log(1-D_r(G(z))) - \sum_{k=1}^{K}(\frac{1}{K}log(D_c(c_k|G(z))+ (1-\frac{1}{K})log(1-D_c(c_k|G(z)))],$$





[ELEB., ICCC, 2017]

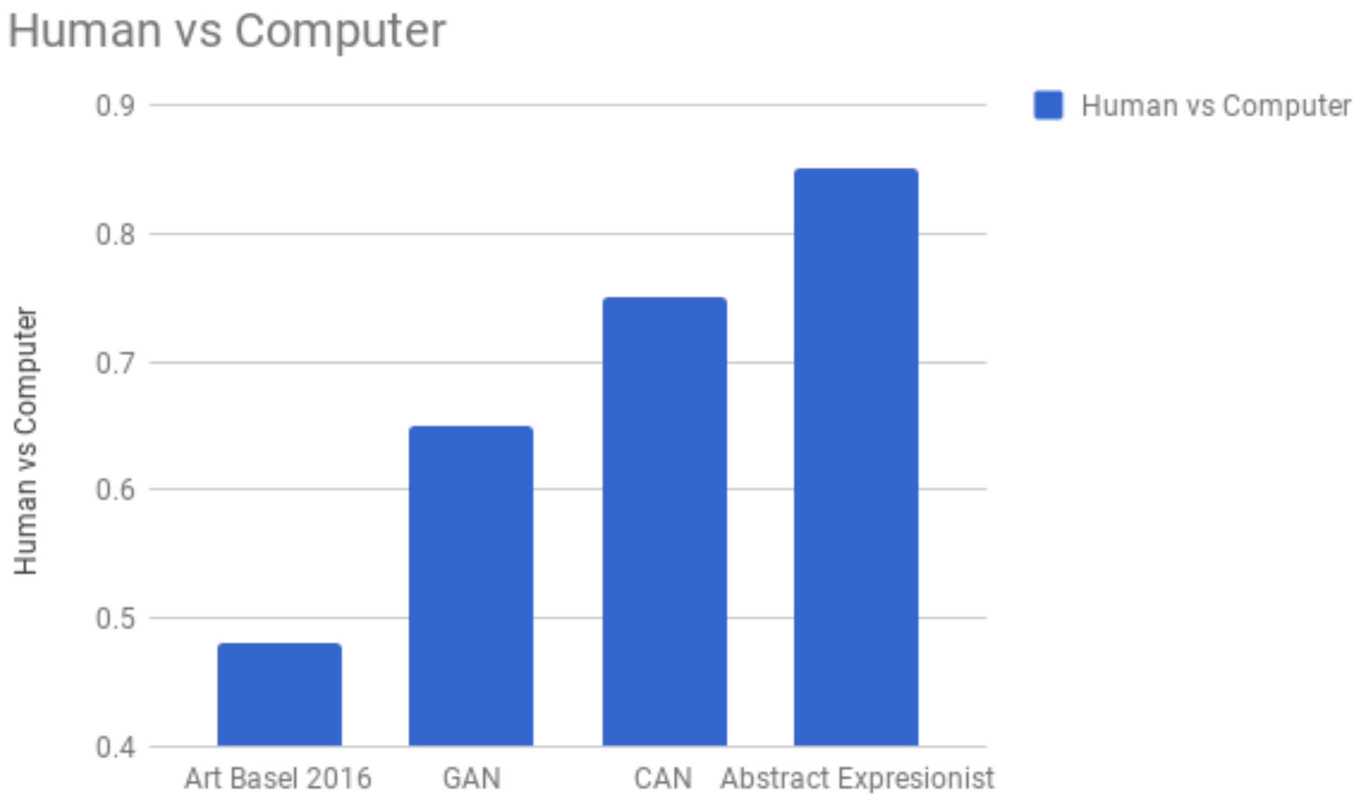
Creative Adversarial Networks

Qualitative Examples

[ELEB., ICCC, 2017]

Creative Adversarial Networks

Human Subject Experiments: Turing Test (Human vs Computer) ~100 images for each set



[ELEB., ICCC, 2017]

Creative Adversarial Networks

Q1: Intentionality

Q2: Structure

Q2: Communication

Q4: Inspiration

Q1: INT Q2:STR Q3:COM Q4:INS

Painting set	Q1 (std)	Q2 (std)	Q3 (std)	Q4 (std)
CAN	3.3 (0.47)	3.2 (0.47)	2.7 (0.46)	2.5 (0.41)
Abstract Expressionist	2.8 (0.43)	2.6 (0.35)	2.4 (0.41)	2.3 (0.27)
Art Basel 2016	2.5 (0.72)	2.4 (0.64)	2.1 (0.59)	1.9(0.54)
Artist sets combined	2.7 (0.6)	2.5 (0.52)	2.2 (0.54)	2.1 (0.45)

[ELEB., ICCC, 2017]

CAN Impact

In addition to the scientific impact, CAN has also been covered at

- Media attention:
  - MIT tech review,
  - New scientist
  - Others
- Exhibitions:
  - Frankfurt Book Fair
  - Los Angeles Art Exhibition
- Invited Talks:
  - Best of AI meeting
  - NIPS 2017 Creativity workshop
- FB CAN Demo
  - NIPS17 FB booth
  - FAIR video

[ELEB., ICCC, 2017]

HBO Silicon Valley TV Series

Creative Fashion Networks

[JACKETS & SWEATERS]

[Sbai*, Elhoseiny*, Bordes, Lecun, Couprie, 2018]

Creative Fashion Networks

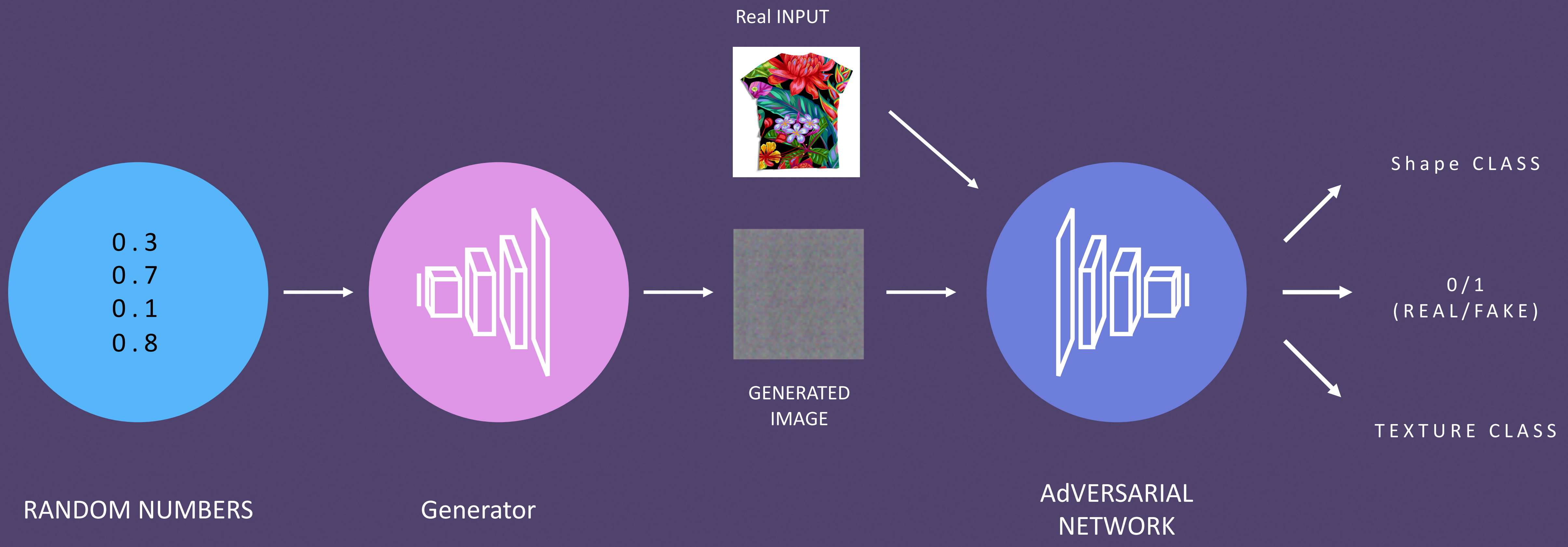
AI Creativity Potential impact

- benefit people's experience

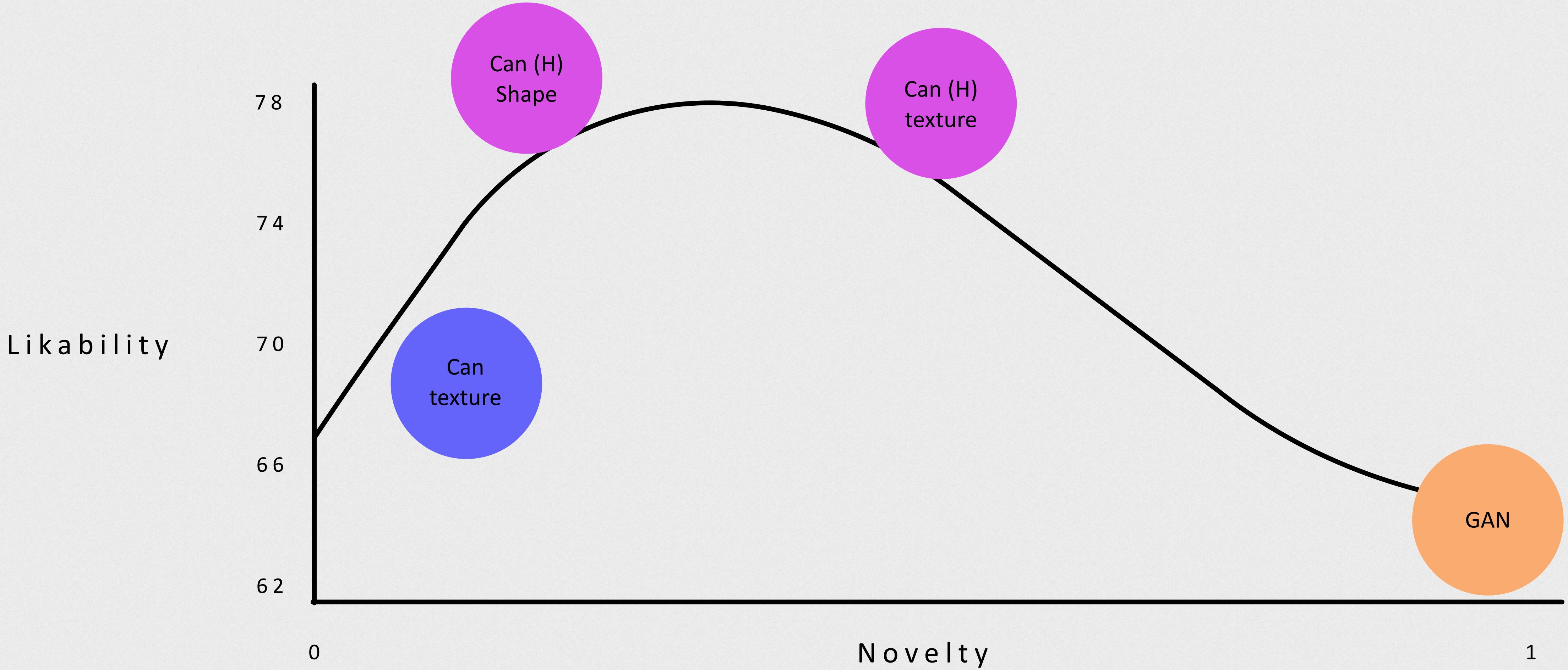
- Additional sources of inspiration for creating unexpected products that are related to the brand DNA.

[S*E*BLC, 2018]

Creative Fashion Networks



Creative Models are Most Popular



judged by humans and measured as a distance to similar training images

[SECBL, 2018]

"interesting" Shapes

[Slide credit, our F8 presentation with Camille Couprie]

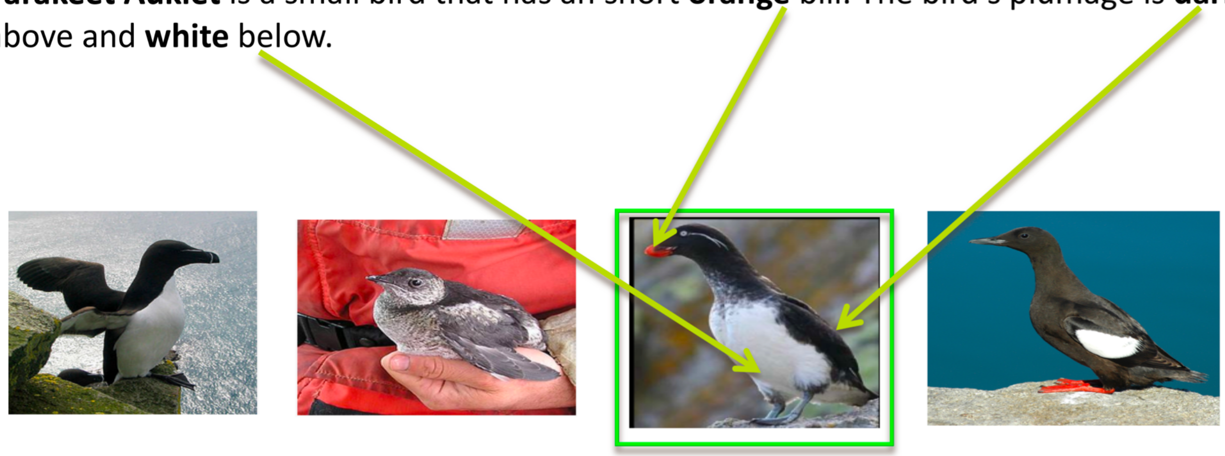
Creative Fashion Generation

- ECCV18 workshop best paper award
- Media attention
  - New scientist
- F8 conference presentation
  - High impact main Facebook conference

Creativity/Ambiguity Loss loops back to help understanding the unseen

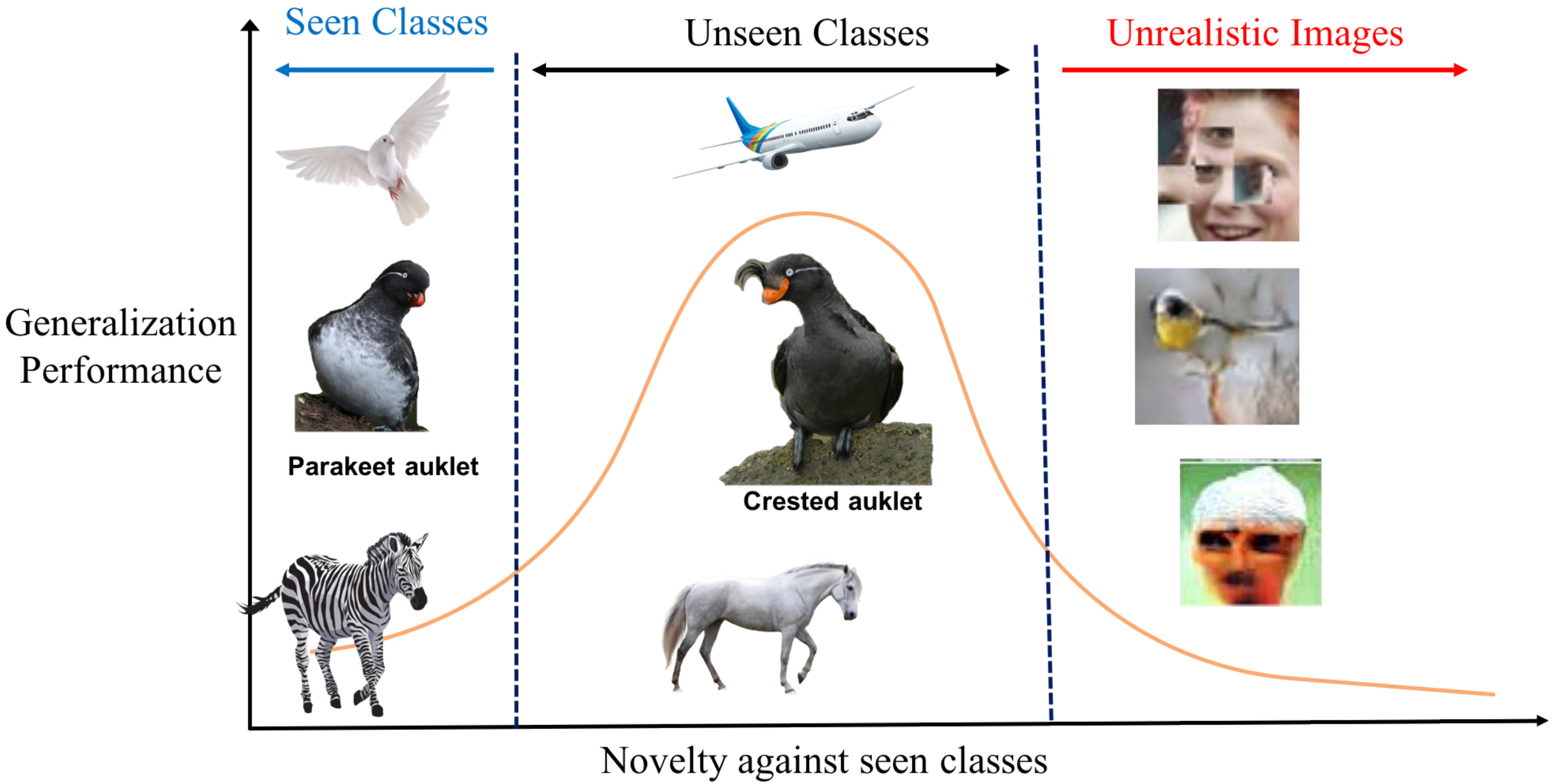
IMAGINE TO SEE

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

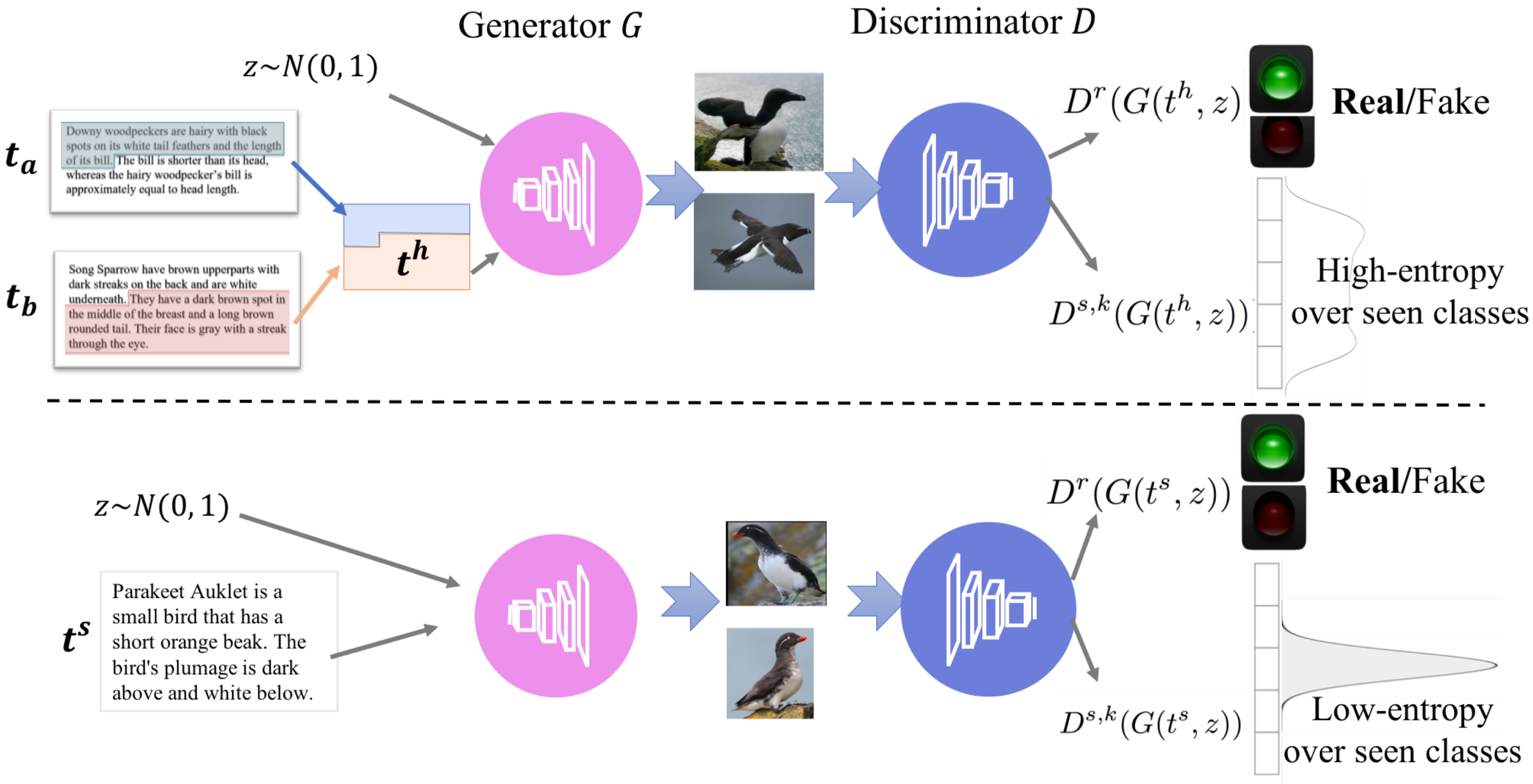


IMAGINE TO CREATE

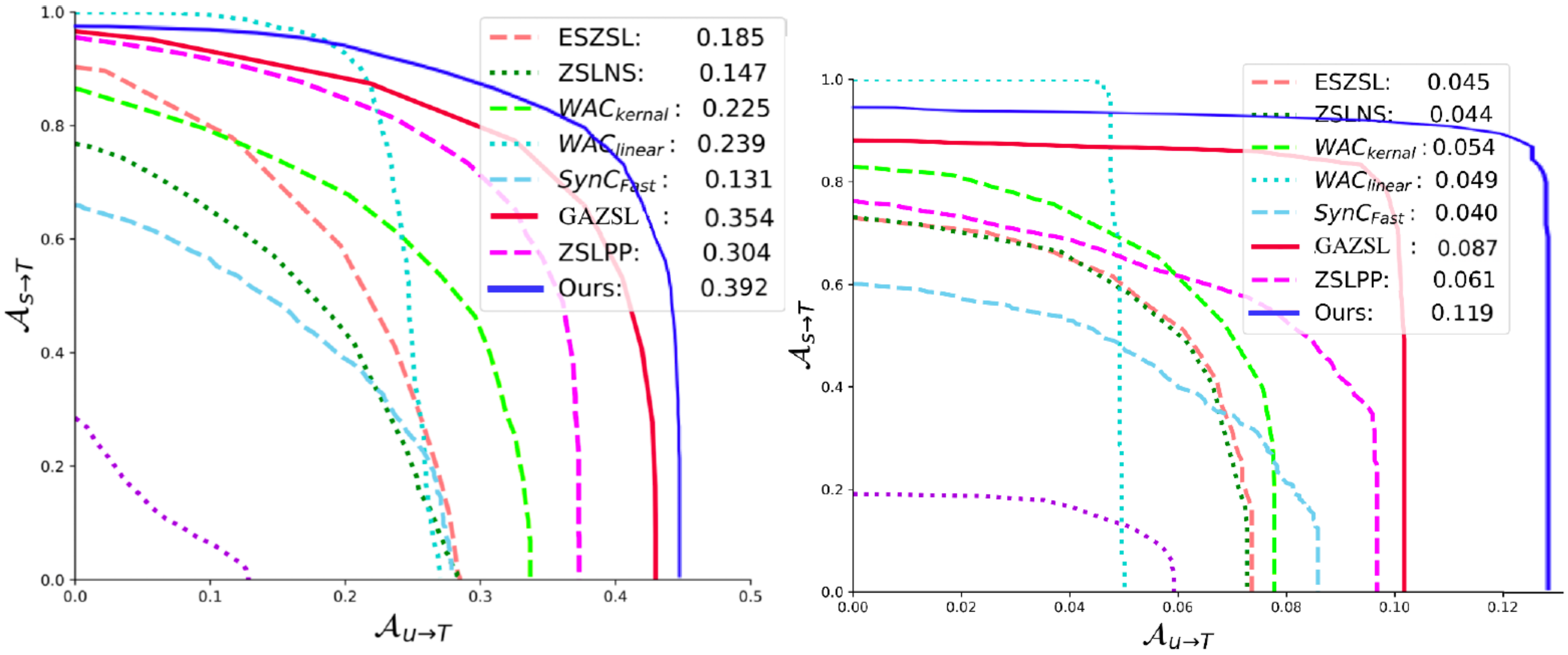
Creativity Inspired Zero-Shot Learning, submitted



Creativity Inspired Zero-Shot Learning, submitted

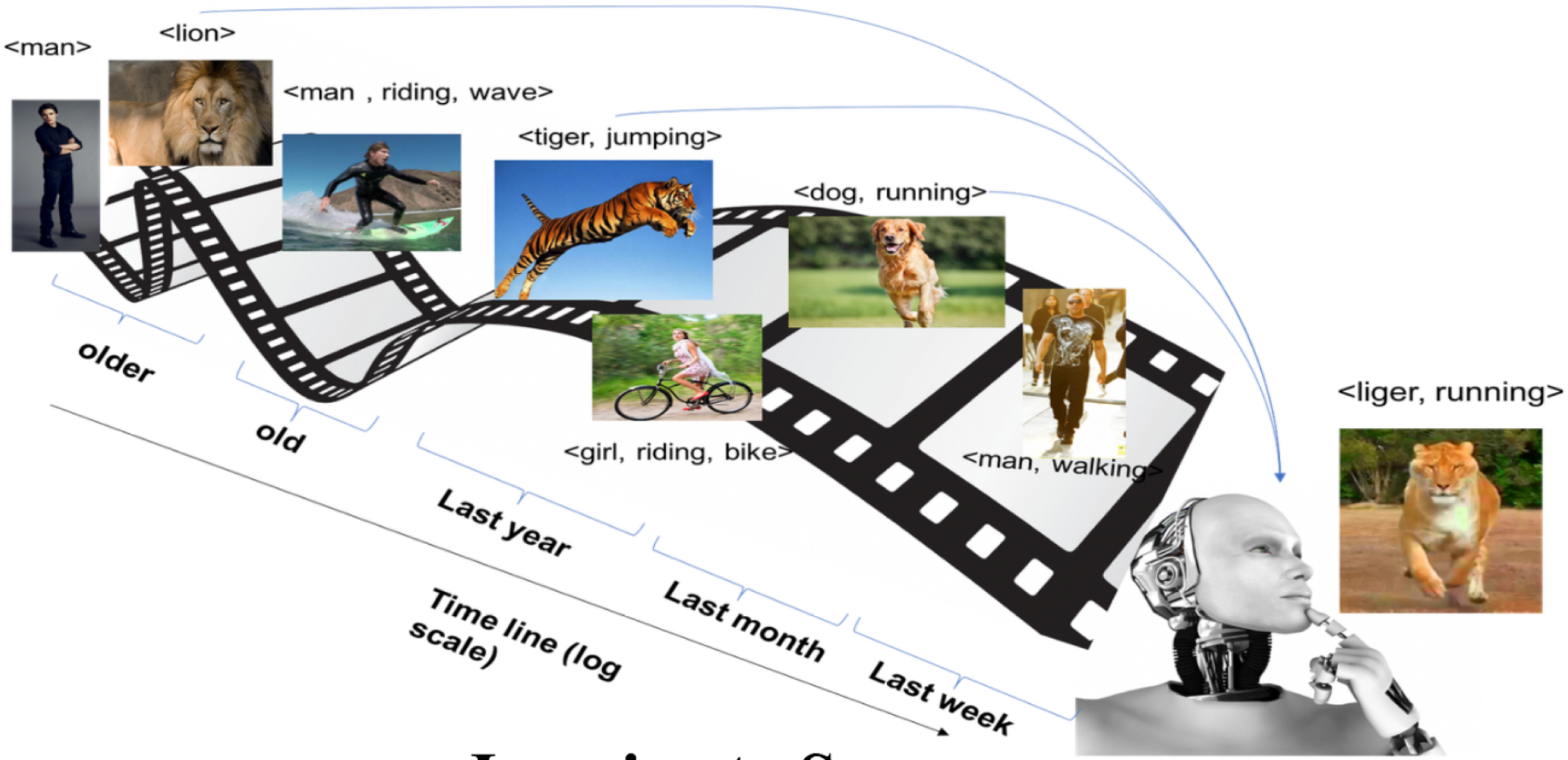


Generalized ZSL Results on CUB



(a) CUB with SCS splitting (b) CUB with SCE splitting

Ongoing Work



Imagine to See

KAUST
King Abdullah University of Science and Technology

Baidu Research

Earlier work

[1] Large Scale Visual Relationship Understanding, [Zhang, Khaldis, Paluri, Rohbrach, Elhoseiny , AAAI, 2019

References

2019 Submissions/Publications(2 publications and 3 preprints, 4/5 as a main FB author)

1. **M Elhoseiny,** M Elfeki, **"Creativity Inspired Zero Shot Learning", CVPR submission, 2019**
2. **M. Elfeki, C. Couprie, M. Elhoseiny,** "GDPP: Learning Diverse Generations using Determinantal Point Processes", **ICML** S, 2019
3. S Ebrahimi, **M Elhoseiny,** T Darrell, **M Rohrbach**, "Uncertainty-guided Lifelong Learning in Bayesian Networks", **ICML** S, 2019
4. **A. Chaudhry, M. Ranzato, M Rohrbach**, **M. Elhoseiny,** "Efficient Lifelong Learning with A-GEM", **ICLR** , 2019
5. **J. Zhang,Y. Khaladis, M. Rohbrach, M. Paluri, M. Elhoseiny,** "Large-Scale Visual Relationship Understanding", **AAAI**, 2019 (**accepted**)

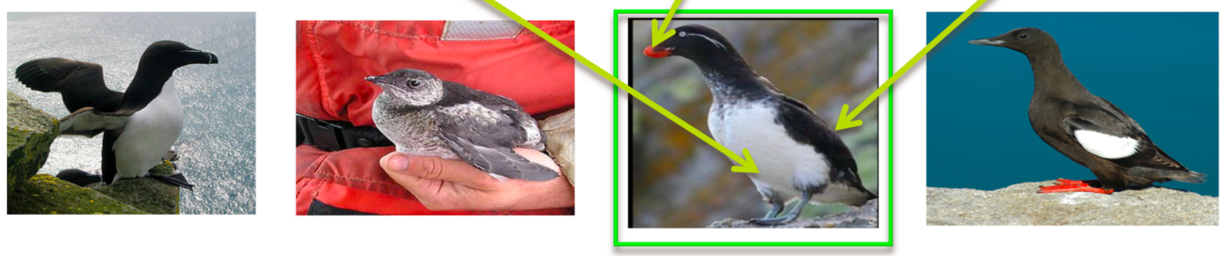
2018 Publications (5-6/7 publications as a main FB author)

1. R Aljundi, F Babiloni, **M Elhoseiny, M Rohrbach**, T Tuytelaars, "Memory Aware Synapses: Learning what (not) to forget", **ECCV** 2018
2. **R. Selvaraju,** P. Chattopadhyay, **M Elhoseiny, T Sharma, D. Batra, D. Parikh**, S. Lee, "Choose your Neuron: Incorporating Domain Knowledge through Neuron Importance", **ECCV**, 2018
3. Y. Zhu, **M. Elhoseiny**, B Liu, A. Elgammal, "Generative Adversarial Approach for Zero-Shot Learning from Noisy Texts", **CVPR**, 2018
4. **O. Sbai*, M. Elhoseiny*, C. Couprie, A. Bordes, Y. LeCun**, "DesIGN: Design Inspiration from Generative Networks", **ECCVW**, 2018, **best paper award, also JMLR 19 submission**
5. A. Elgammal, M. Mazzone, , B. Liu, and D. Kim, **M. Elhoseiny,** and "The Shape of Art History in the Eyes of the Machine", **AAAI**, 2018 **(oral)**
6. **M Elhoseiny,** F Babiloni, R Aljundi**, M Rohrbach**, T Tuytelaars, "Towards Human-Like Life-long Fact Learning", **ACCV**, 2018
7. **M. Elhoseiny**, Yi. Zhu, and Ahmed Elgammal**, "Language Guided Visual Recognition", Deep Learning Semantic Recognition Book,** 2018

Thank you, Questions?

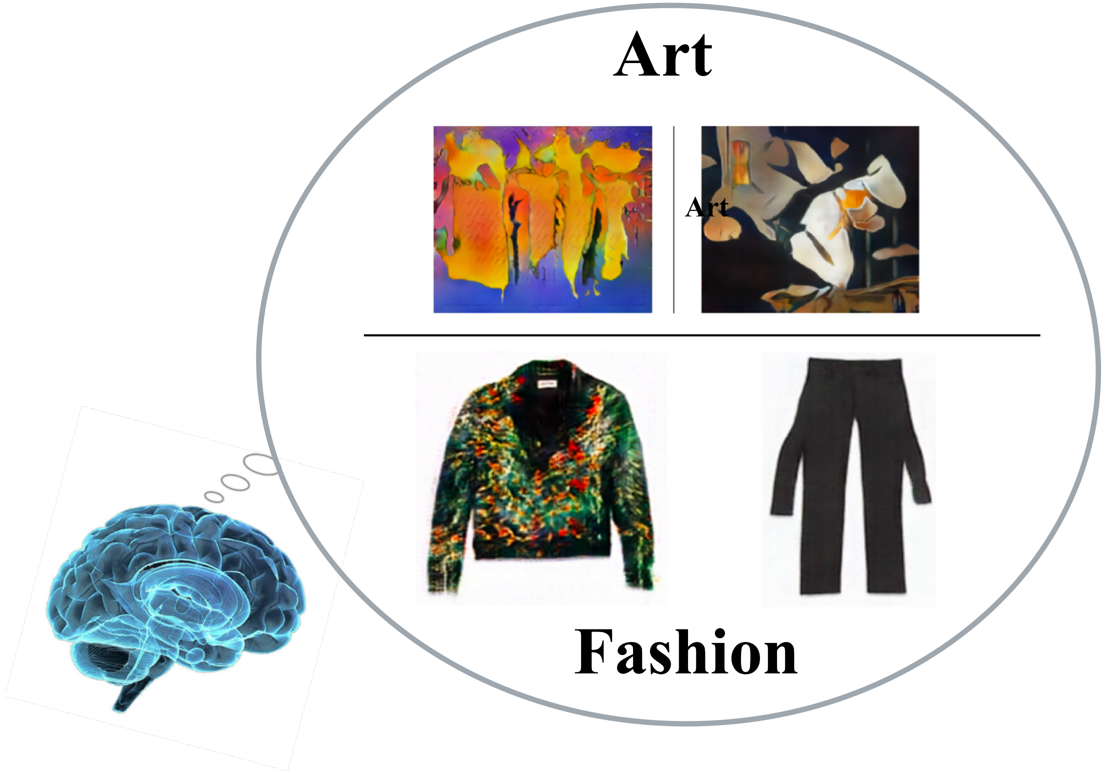
IMAGINE TO SEE

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



[ESE, ICCV, 2013],
[EES, TPAMI, 2016]
[EZE, CVPR, 2017]
[ZELE,CVPR, 2018]

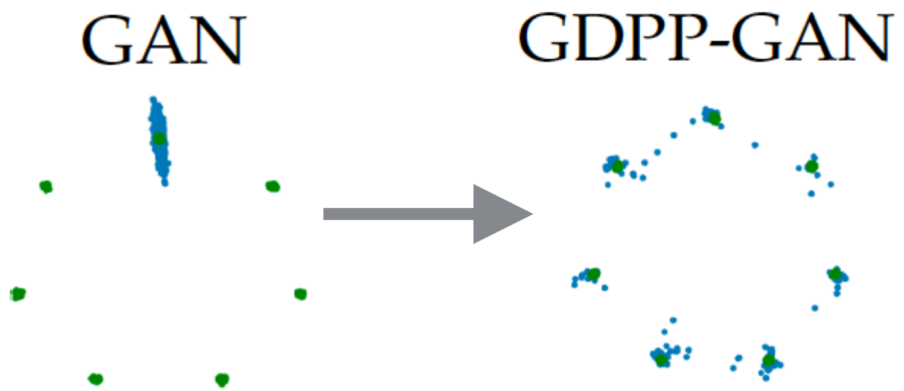
IMAGINE TO CREATE



[ELEB., ICCC, 2017]
[SEBLC, 2018]

mohamed.elhoseiny@kaust.edu.sa

Helping the Imaginer:
Generative DPP, ICML19 Submission

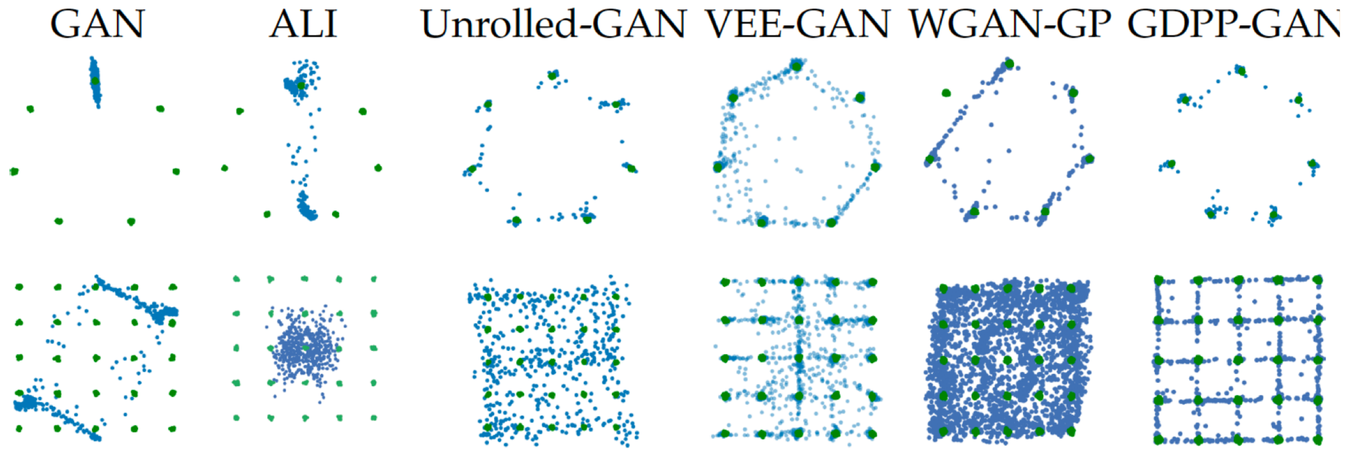


Our loss only requires a generator G and a feature extraction function ϕ, which is:

1. Resisting mode collapse
2. Data Efficient
3. Time Efficient
4. Stabilizes adversarial training
5. Producing higher quality samples
6. Architecture invariant
7. Unsupervised: No labels
8. Cost free: No trainable parameters
9. Generic: The loss can be added to *ANY* generative model.

[Elfeki, Couprie, Elhoseiny, ICLR 2019 submission]

Helping the Imaginer:
Generative DPP with M Elfeki and C. Couprie, ICML19 S



	2D Ring		2D Grid		1200D Synthetic	
	Modes (Max 8)	% High Quality Samples	Modes (Max 25)	% High Quality Samples	Modes (Max 10)	% High Quality Samples
GAN	1	99.3	3.3	0.5	1.6	2.0
ALI	2.8	0.13	15.8	1.6	3	5.4
Unrolled GAN	7.6	35.6	23.6	16.0	0	0.0
VEE-GAN	**8.0**	52.9	24.6	40.0	5.5	28.3
WGAN-GP	6.8	59.6	24.2	28.7	6.4	29.5
GDPP-GAN	**8.0**	**71.7**	**24.8**	**68.5**	**7.4**	**48.3**

Data and Time Efficiency

2D-Grid: Data & Time Efficiency