# Neural Networks for Natural Language Processing

Jun.-Prof. Dr. Sophie Fellenz

## Week 12 – Self-Supervised Learning – Style Transfer

27 Jan 2025

Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

# Course Organization

- Scheduling of Q&A Session
- Last Exercise Sheet due today



RPTU

# Outline Self-Supervised

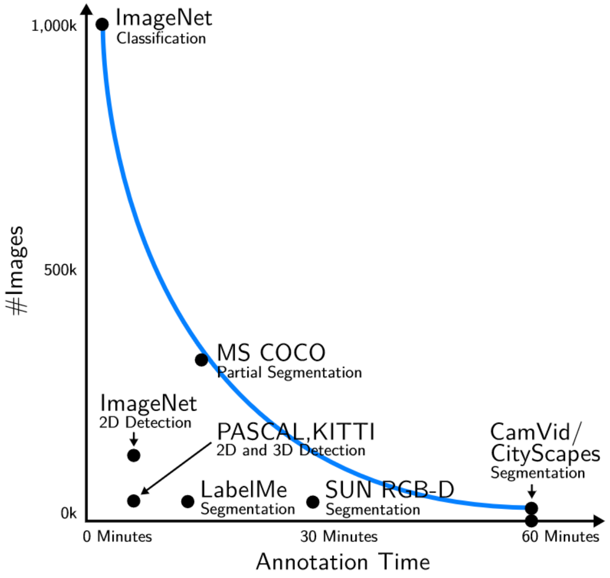
- Preliminaries
- Pretext Tasks
- Self-Supervised Learning Concepts
- Contrastive Learning
    - Quick-Thoughts
    - CLIP
- Summary

# What is self-supervision?

https://t3.ftcdn.net/jpg/03/12/24/14/360_F_312241475_OywzPQNBkO4xkSpT9vYDuJZyppn37aHM.jpg

https://ais-akamai.rtl.de/masters/1373178/1686x0/babyentwicklung-baby-spielt-mit-bunten-ringen.jpg

# Why self-supervision?

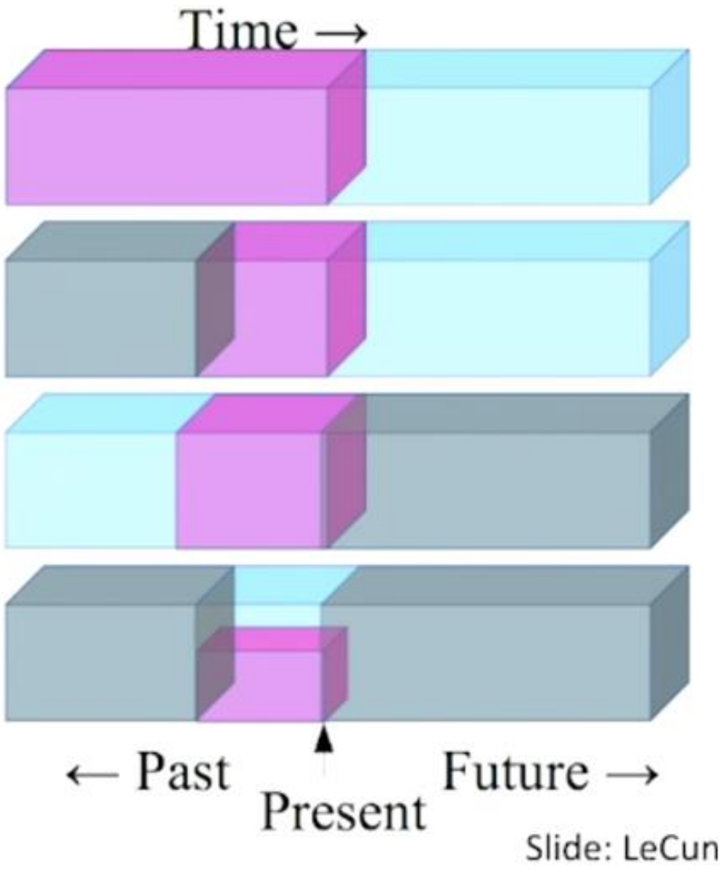



https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/autonomous-vision/lectures/computer-vision/

- Getting labels for supervision is expensive
  - E.g. Labeling Imagenet took 22 human years
- Self-supervision from pseudo-labels for free

# Idea of Self-Supervision

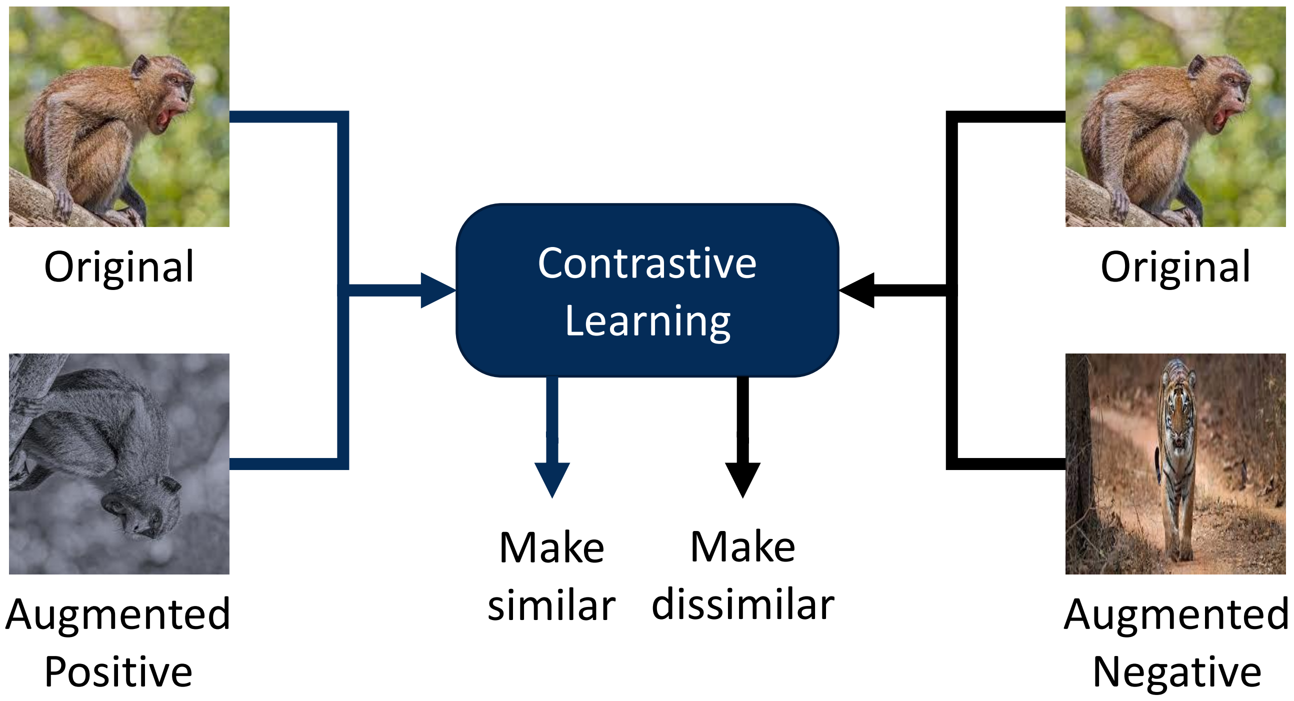
- Predict any part of the input from any other part.
- Predict the future from the past.
- Predict the future from the recent past.
- Predict the past from the present.
- Predict the top from the bottom.
- Predict the occluded from the visible
- Pretend there is a part of the input you don't know and predict that.




Slide credits: Yann LeCun and Ishan Misra



RPTU




Jaiswal 2020, https://images.app.goo.gl/cehS4AmHg1tvjzcF9;
https://images.app.goo.gl/u8UthNzMf7ooyDj86



RPTU

# Learning problems

- Unsupervised learning
  - Learn model parameters using data without labels $\{\mathbf{x_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$
  - Examples: Clustering, dimensionality reduction, generative models
- Supervised learning
  - Learn model parameters using data with labels $\{(\mathbf{x_i}, \boldsymbol{y_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$
  - Examples: Classification, regression
- Self-supervised learning
  - Learn model parameters using data-data pairs $\{(\mathbf{x_i}, \boldsymbol{x_i}')\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$
  - Examples: Contrastive learning

# Pretext task to learn representations

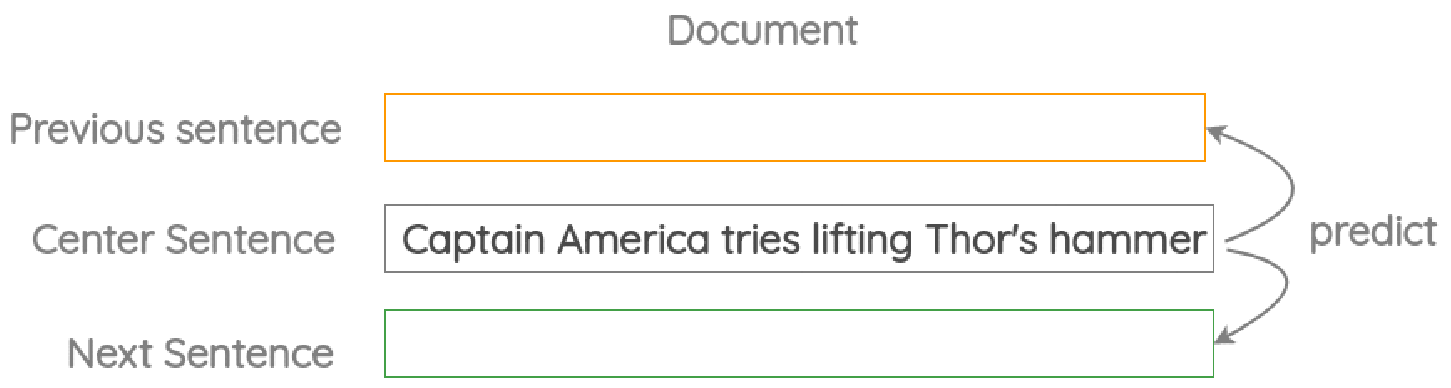

- Learn more general representations using self-supervision
- Use lots of unlabeled data for pre-training on pretext task
- Pretext Task ≠ Target Task

# Skip-Gram


- Goal: Predict context words from center word
- Example
  - Context size 1
  - Predict 2 surrounding words from center word

https://amitness.com/2020/05/self-supervised-learning-nlp/
[Mikolov et al. 2013]

# Skip-Thoughts



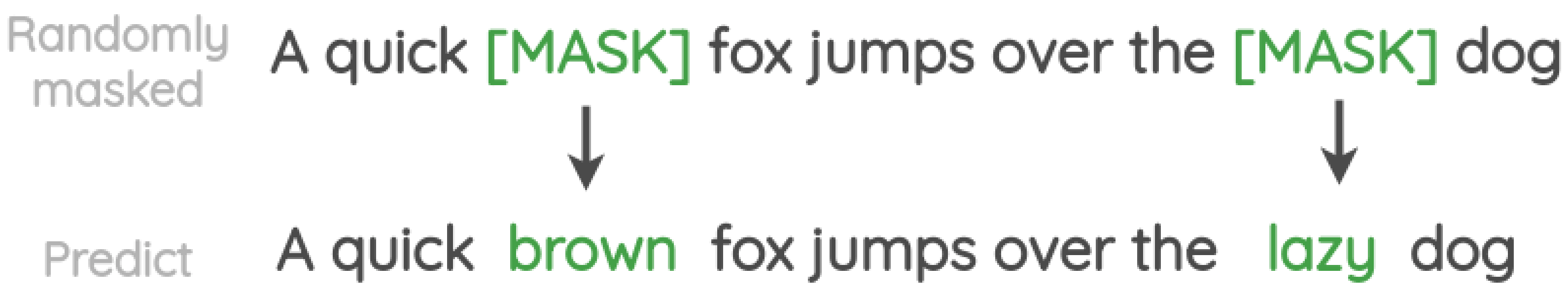

- Goal: Predict neighboring sentences
- Example
  - Context size 1
  - Predict 2 surrounding sentences from center sentence

https://amitness.com/2020/05/self-supervised-learning-nlp/
[Kiros et al. 2015]



RPTU

# Masked language model



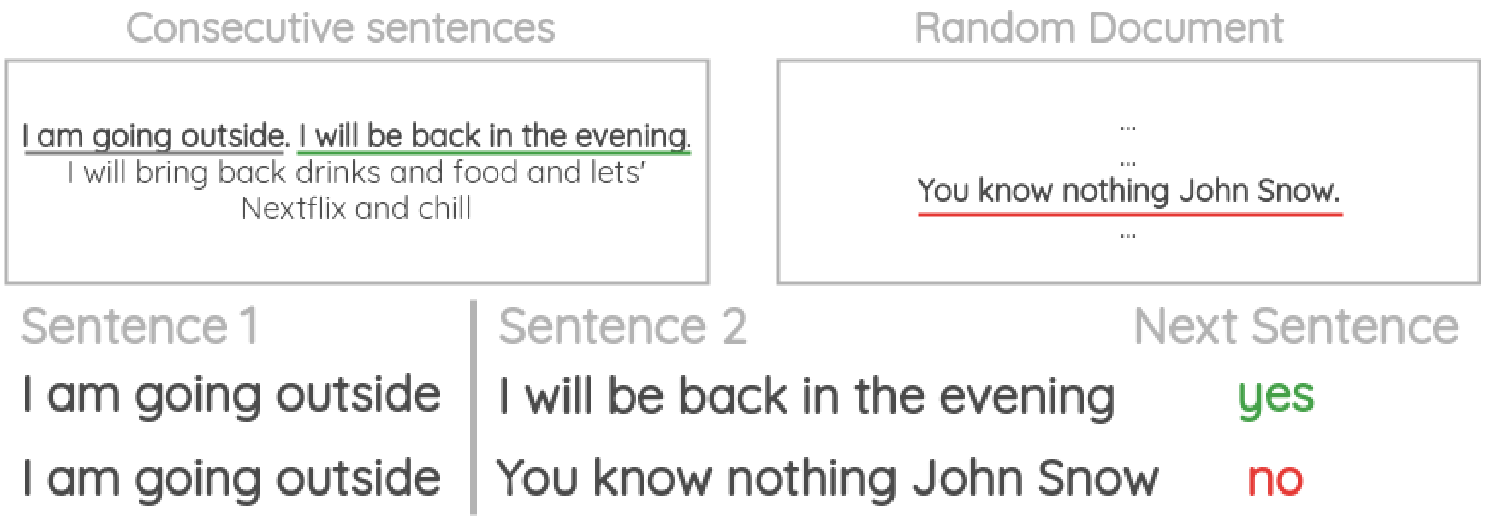

- Randomly mask text
- Model predicts masked text from surrounding words
- Used in combination with next sentence prediction for pre-training BERT

https://amitness.com/2020/05/self-supervised-learning-nlp/
[Devlin et al. 2019]



RPTU

# Next sentence prediction

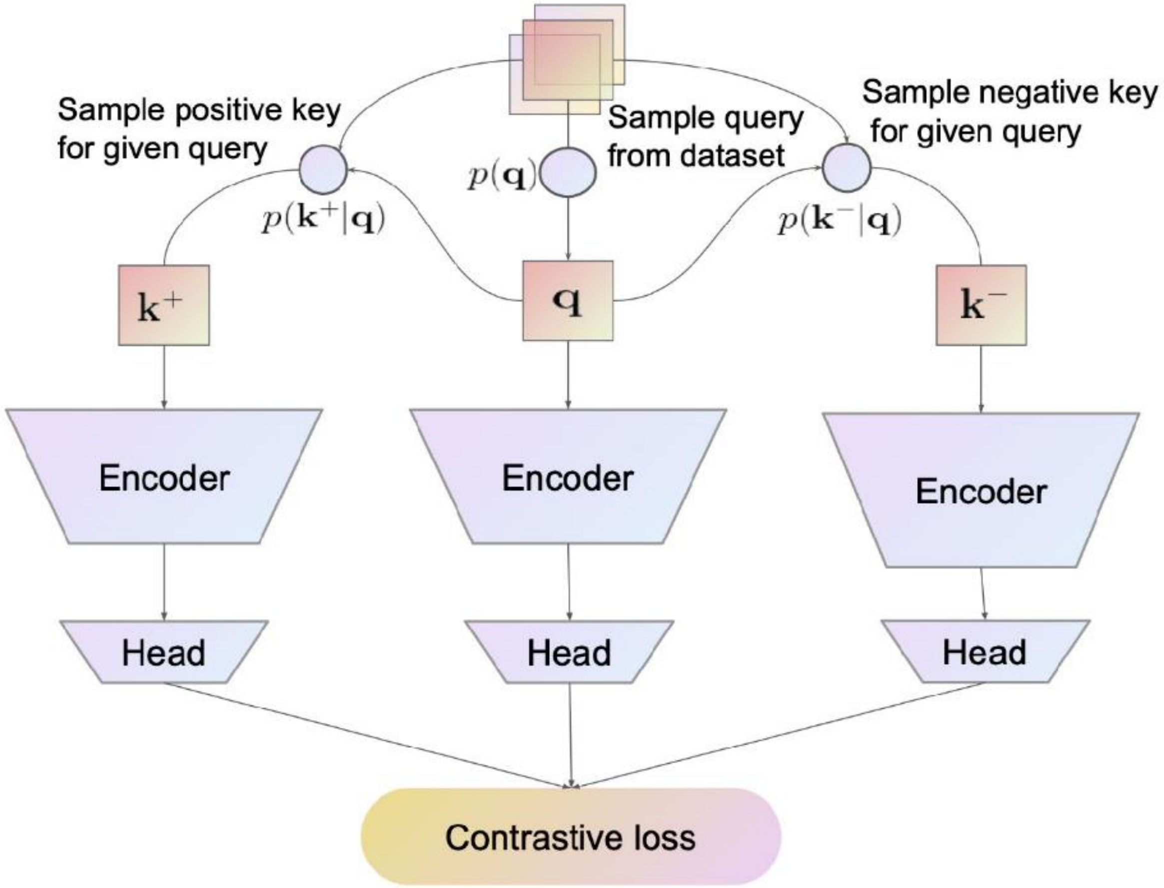



https://amitness.com/2020/05/self-supervised-learning-nlp/
[Devlin et al. 2019]

# Pretext Tasks in NLP

- Generative
  - Auto-regressive language modeling
  - Continuous Bag of Words, Skip-Gram
  - Skip-Thoughts
  - Masked language model, next sentence prediction
- Contrastive
  - Quick-Thoughts
  - MI Maximization
- Generative-Contrastive
  - Replaced token detection

# Contrastive loss




[Le-Khac et al. 2020]



RPTU

# Contrastive losses

- Traditional losses
  - Discriminative models measure losses with respect to prediction label
  - Generative models measure losses in the input space
- Contrastive losses
  - Target is defined in terms of metric embeddings instead of fixed targets
  - Loss measured in embedding space
  - Decomposed into scoring functions and the actual form

[Le-Khac et al. 2020]

# Contrastive learning objective - similarity

$$\mathcal{L} = \mathbb{E}_{x,x^+,x^-}\left[-\log\left(\frac{\exp(sim(f(x),f(x^+))}{\exp(sim(f(x),f(x^+)) + \exp(sim\big(f(x),f(x^-)\big)}\right)\right]$$

- Similarity functions
  - Distance: Euclidean

$$sim(x,y) = \|x-y\|_2$$

  - Similarity: Inner product or (normalized) cosine similarity

$$sim(x,y) = \|f(x)^T f(y)\|_2$$



RPTU

# Noise Contrastive Estimation

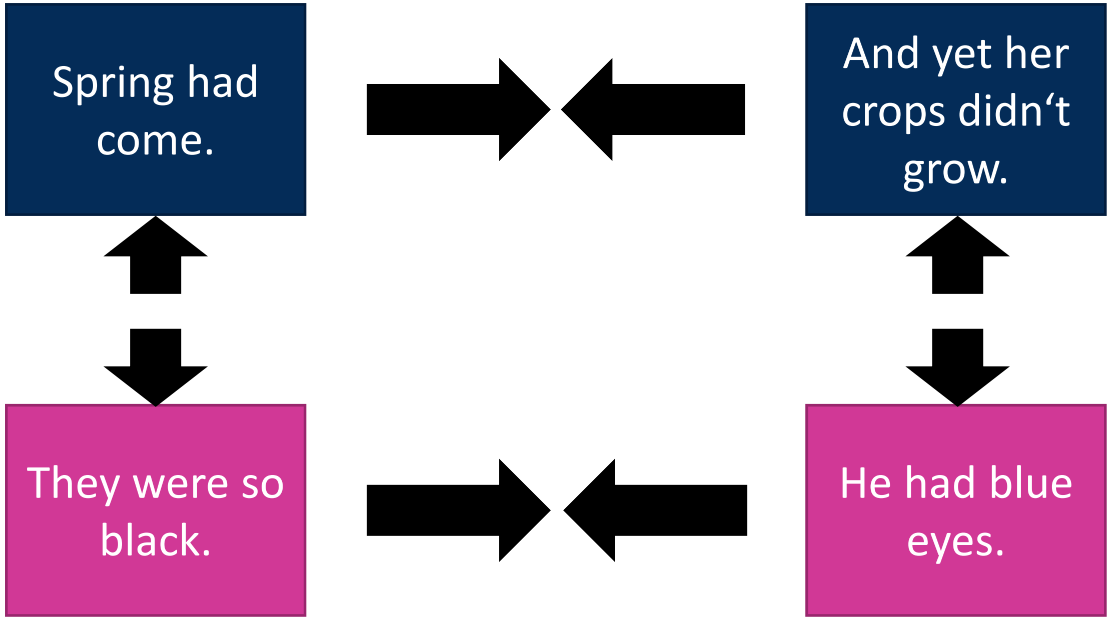
$$\mathcal{L} = \mathbb{E}_{x,x^+,x^-}\left[-\log\left(\frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + e^{f(x)^T f(x^-)}}\right)\right]$$

- Encoder f and similarity measure (here inner product) may be exchanged based on the task, framework stays the same
- For more negative samples: InfoNCE

$$\mathcal{L} = \mathbb{E}_{x,x^+,x^k}\left[-\log\left(\frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + \sum_{k=1}^{K} e^{f(x)^T f(x^k)}}\right)\right]$$

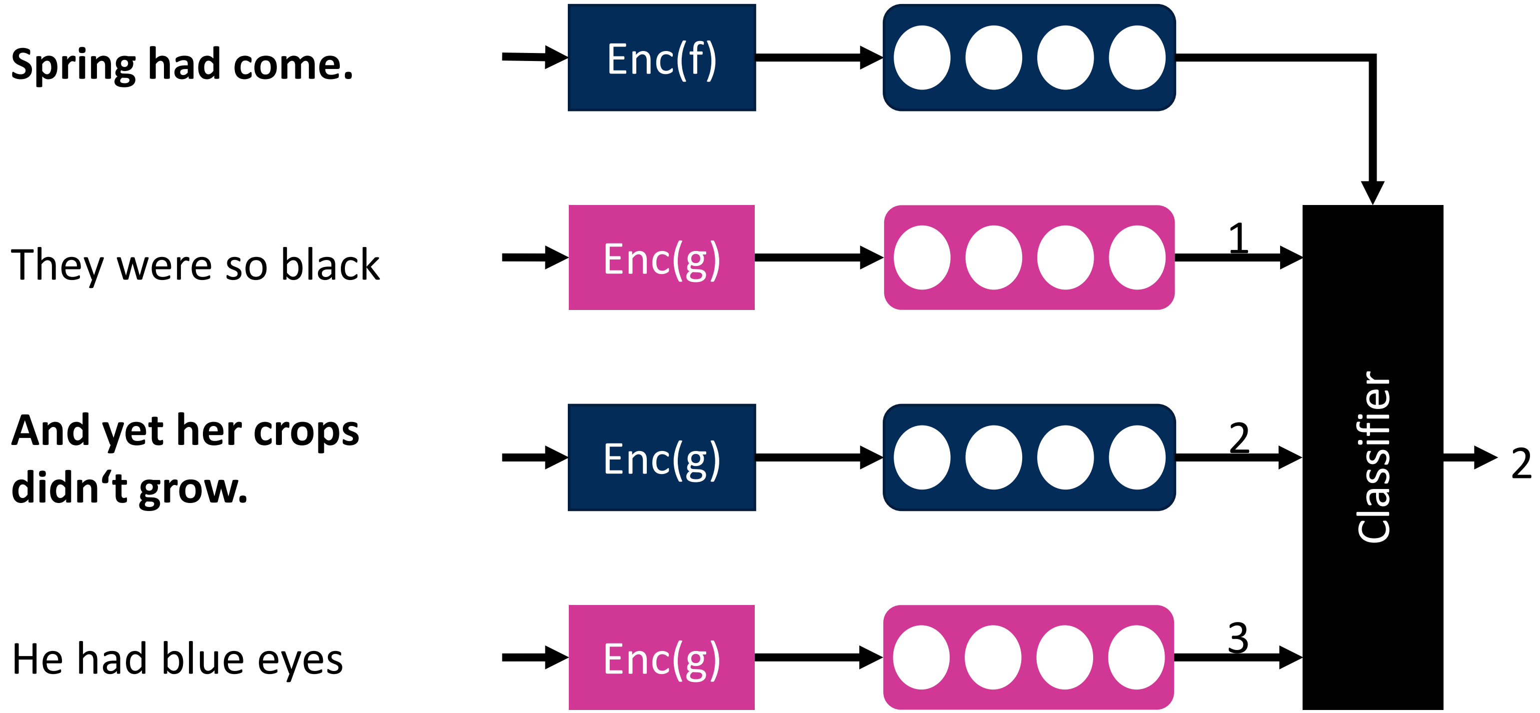
[Gutmann et al. 2010, Oord et al. 2018]

# Quick-Thoughts basic idea




[Logeswaran et al. 2018]

# Quick-Thoughts basic architecture




[Logeswaran et al. 2018]

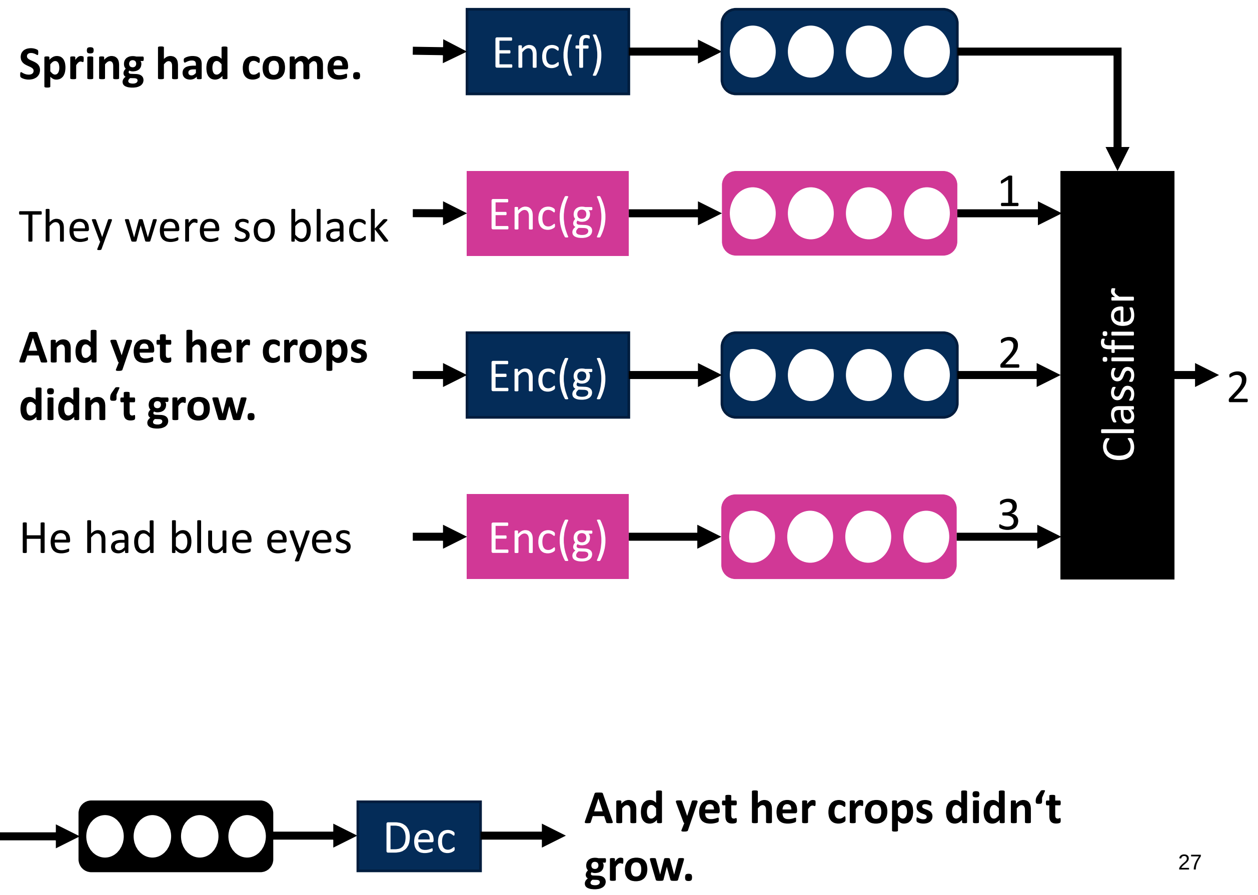


RPTU

Quick-Thoughts

vs

Skip-Thoughts




27



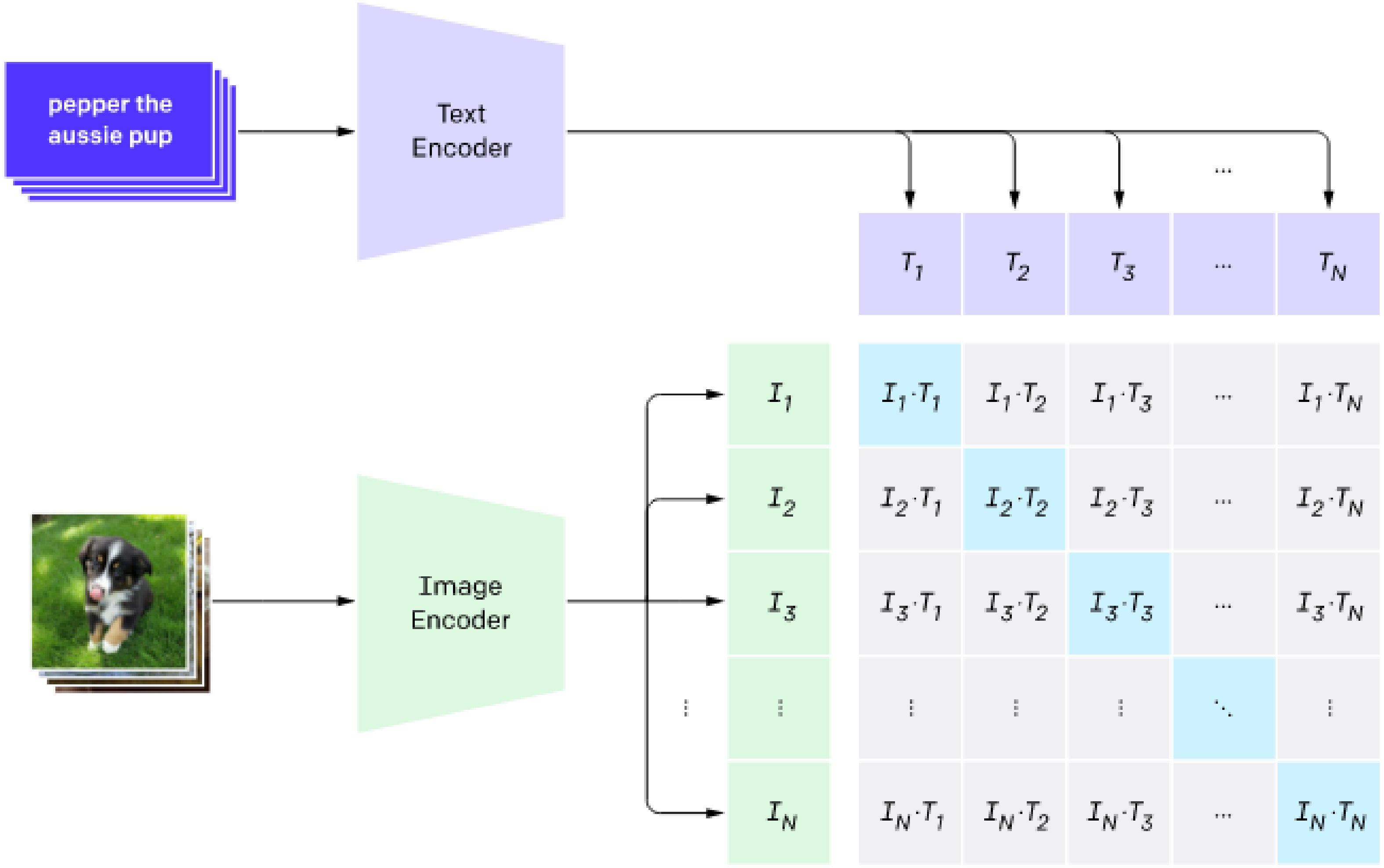
RPTU

# CLIP – Contrastive Language-Image Pre-Training

- Learns to associate images and natural language by connecting visual concepts with natural language supervision
- Dataset is created from abundance of image-caption pairs from the internet

[Radford et al. 2021]

# CLIP - Pre-training

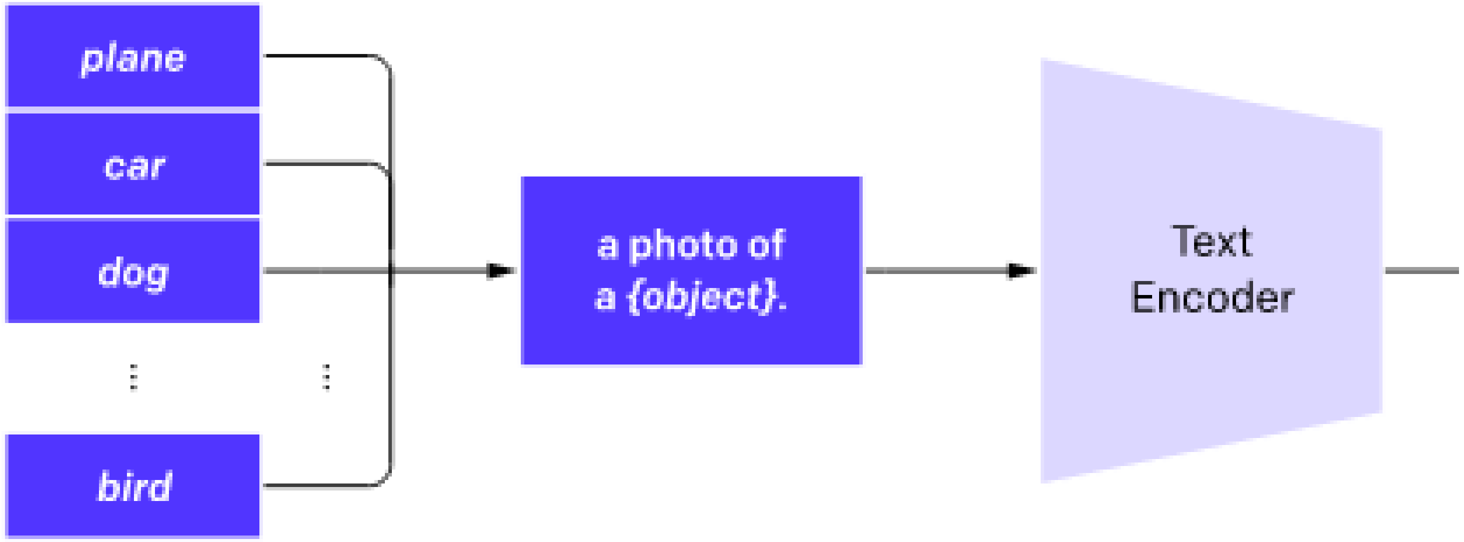





RPTU

# CLIP

Transfer dataset labels to common format

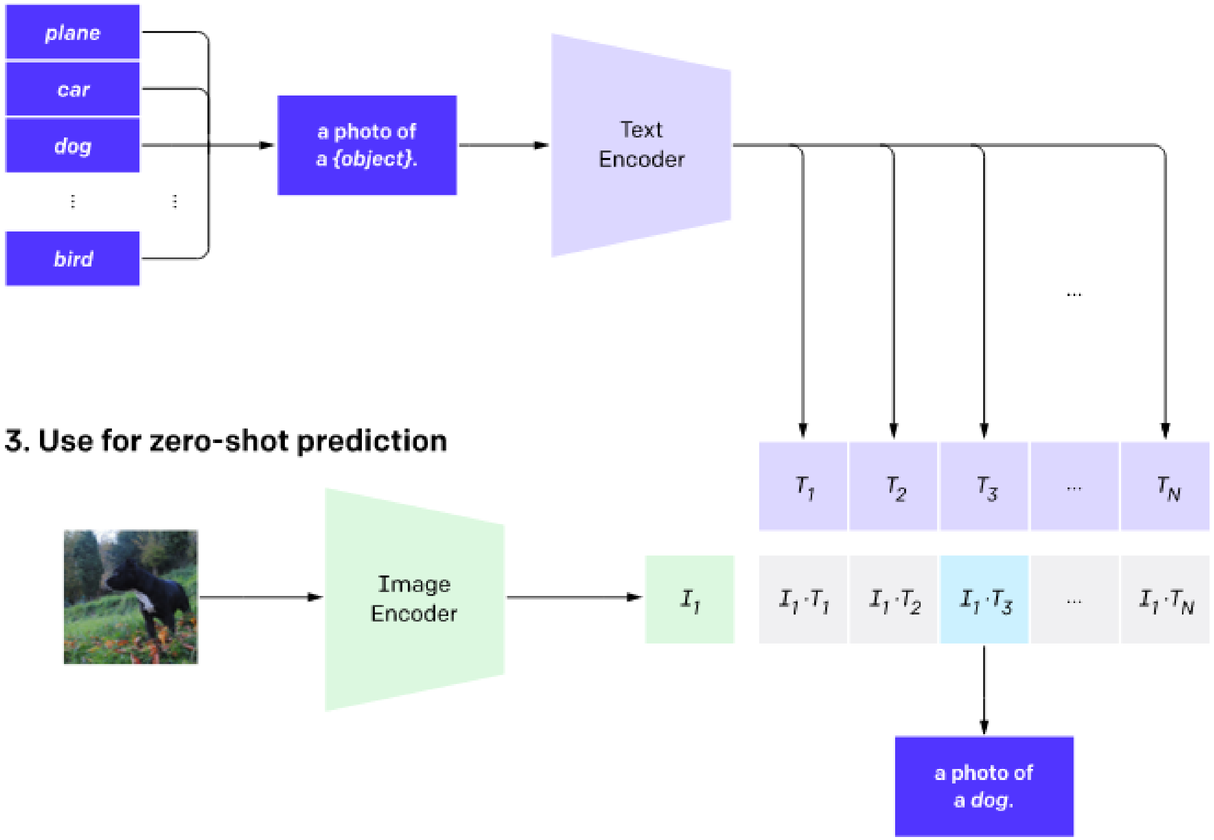





RPTU

# CLIP

Use transferred dataset labels to create classifier for zero-shot prediction






RPTU

# CLIP performance

| DATASET | IMAGENET RESNET101 | CLIP VIT-L |
|---|---|---|
|  ImageNet | 76.2% | 76.2% |
|  ImageNet V2 | 64.3% | 70.1% |
|  ImageNet Rendition | 37.7% | 88.9% |
|  ObjectNet | 32.6% | 72.3% |
|  ImageNet Sketch | 25.2% | 60.2% |
|  ImageNet Adversarial | 2.7% | 77.1% |



RPTU

# CLIP takeaways

**FOOD101**

**guacamole** (90.1%) Ranked 1 out of 101 labels

✓ a photo of **guacamole**, a type of food.

× a photo of **ceviche**, a type of food.

× a photo of **edamame**, a type of food.

× a photo of **tuna tartare**, a type of food.

× a photo of **hummus**, a type of food.

**SUN397**

**television studio** (90.2%) Ranked 1 out of 397

✓ a photo of a **television studio**.

× a photo of a **podium indoor**.

× a photo of a **conference room**.

× a photo of a **lecture room**.

× a photo of a **control room**.

**YOUTUBE-BB**

**airplane, person** (89.0%) Ranked 1 out of 23

✓ a photo of a **airplane**.

× a photo of a **bird**.

× a photo of a **bear**.

× a photo of a **giraffe**.

× a photo of a **car**.

**EUROSAT**

**annual crop land** (12.9%) Ranked 4 out of 10

× a centered satellite photo of **permanent crop land**.

× a centered satellite photo of **pasture land**.

× a centered satellite photo of **highway or road**.

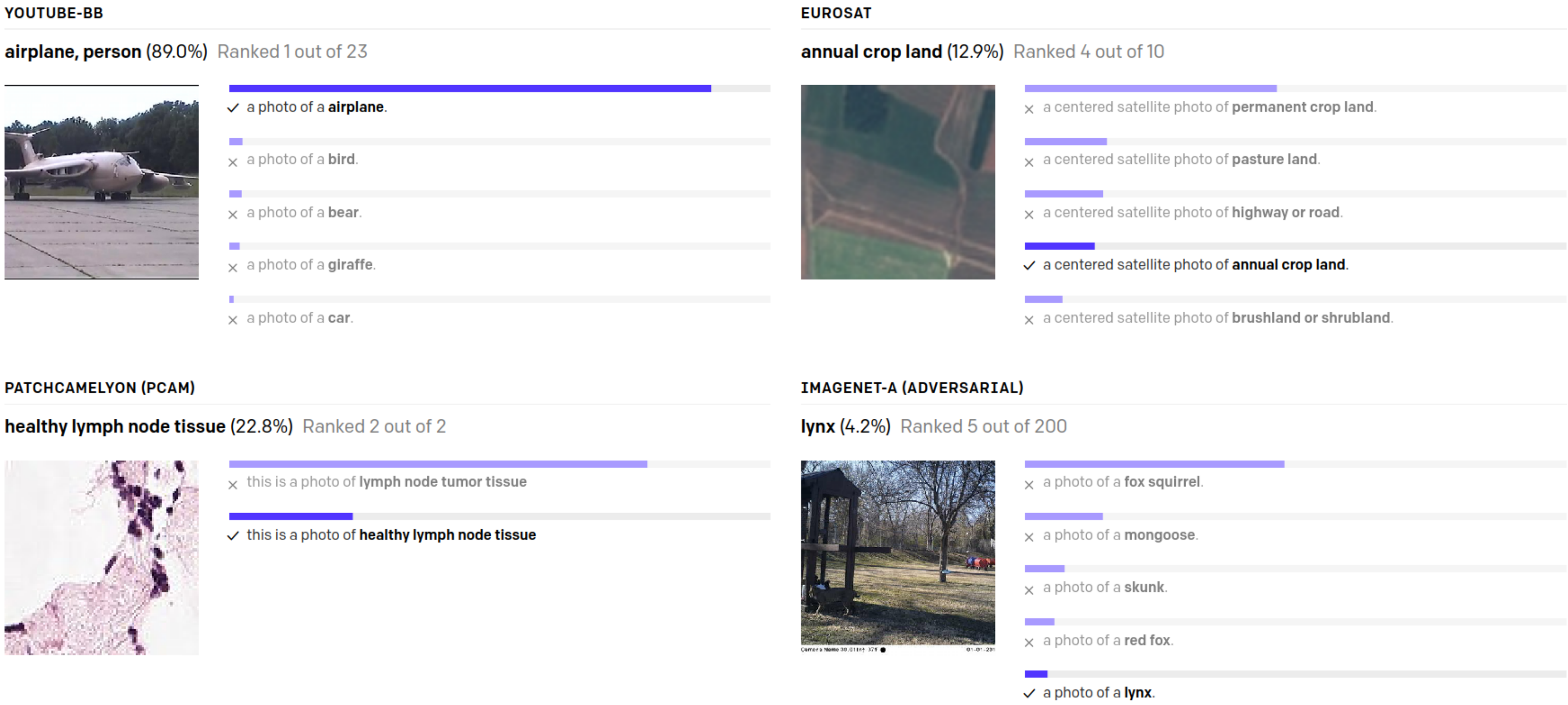
✓ a centered satellite photo of **annual crop land**.

× a centered satellite photo of **brushland or shrubland**.



RPTU

# CLIP takeaways






RPTU

# CLIP objective

- $x_{i,j}$ is the cosine similarity between the i-th image representation $I(p_i)$ and j-th text representation $T(t_j)$
- $y_i$ is the label index
- Overall loss comprises loss term for image-to-text similarity $\mathcal{L}_I$ and text-to-image similarity $\mathcal{L}_{\boldsymbol{T}}$

$$x_{i,j} = \frac{I(p_i) * T(t_j)}{\|I(p_i)\| * \|T(t_j)\|}$$

$$\mathcal{L}_I = -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp(x_{i,y_i})}{\sum_{j=1}^{N}\exp(x_{i,j})} \qquad \mathcal{L}_T = -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp(x_{y_i,i})}{\sum_{j=1}^{N}\exp(x_{j,i})}$$

$$\mathcal{L}_{CLIP} = \frac{\mathcal{L}_I + \mathcal{L}_T}{2}$$

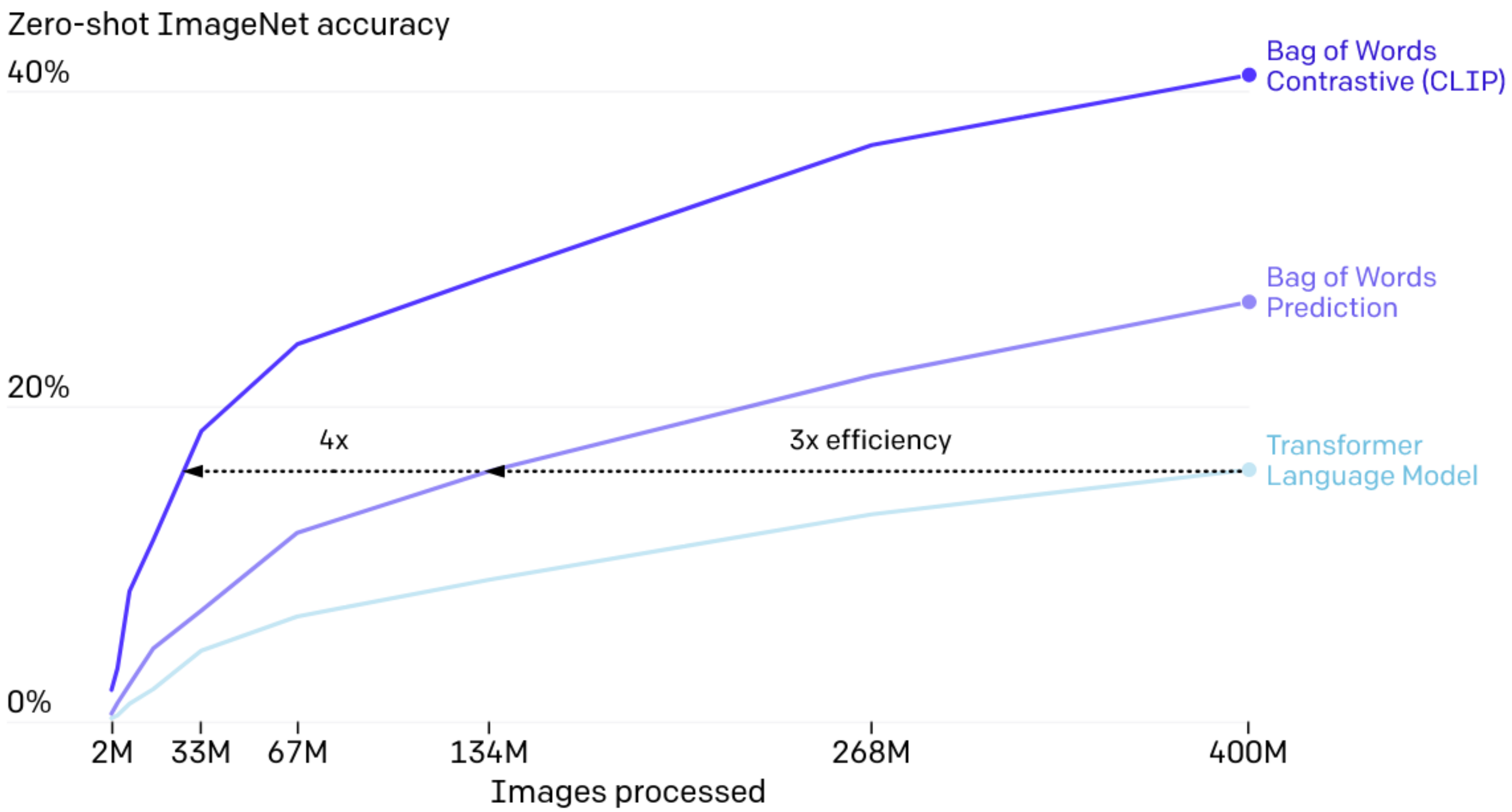
# CLIP code

```
# image_encoder - ResNet or Vision Transformer
# text_encoder  - CBOW or Text Transformer
# I[n, h, w, c] - minibatch of aligned images
# T[n, l]       - minibatch of aligned texts
# W_i[d_i, d_e] - learned proj of image to embed
# W_t[d_t, d_e] - learned proj of text to embed
# t             - learned temperature parameter

# extract feature representations of each modality
I_f = image_encoder(I) #[n, d_i]
T_f = text_encoder(T)  #[n, d_t]

# joint multimodal embedding [n, d_e]
I_e = l2_normalize(np.dot(I_f, W_i), axis=1)
T_e = l2_normalize(np.dot(T_f, W_t), axis=1)

# scaled pairwise cosine similarities [n, n]
logits = np.dot(I_e, T_e.T) * np.exp(t)

# symmetric loss function
labels = np.arange(n)
loss_i = cross_entropy_loss(logits, labels, axis=0)
loss_t = cross_entropy_loss(logits, labels, axis=1)
loss   = (loss_i + loss_t)/2
```



RPTU

# CLIP performance






RPTU

# CLIP takeaways

- Very efficient due to contrastive training objective
- Flexible and general: good zero-shot performance on many tasks
- Prompt engineering is important for good performance
- Poor generalization to anything not covered in the training set
- On fine-grained or abstract classification tasks, task-specific models are still better

# Summary

- Self-Supervised Learning as a workaround for missing labels
- High quality representations from pretext tasks
- NCE as the foundation of contrastive learning
- Contrastive Learning examples
  - Quick-Thoughts for sentence representations
  - CLIP for connecting visual and textual representations

# Text Style Transfer

Rheinland-Pfälzische
Technische Universität
Kaiserslautern
Landau

# Outline

- Adversarial learning (GANs)
- Introduction to text style transfer
- Definition of text style
- Style transfer models
  - Parallel
  - Non-parallel
  - Examples
- Style transfer evaluation

# Adversarial Training

- „Training a model in a worst-case scenario, with inputs chosen by an adversary“
- Examples:
  - An agent playing against a copy of itself in a board game [Samuel, 1959]
  - Robust optimization / robust control [e.g. Rustem and Howe 2002]
  - Training neural networks on adversarial examples [Szegedy et al. 2013, Goodfellow et al. 2014]

# Generative Adversarial Networks

- Both players are neural networks
- Worst case input for one network is produced by another network
- Goal: Generate new samples that look like examples from the training dataset

# GANs


Cop (discriminator)
Tries to distinguish real from fake profiles

Cyber criminal (generator)
Attempts to create online identities that resemble ordinary citizens

# Generative Adversarial Nets (GANs)

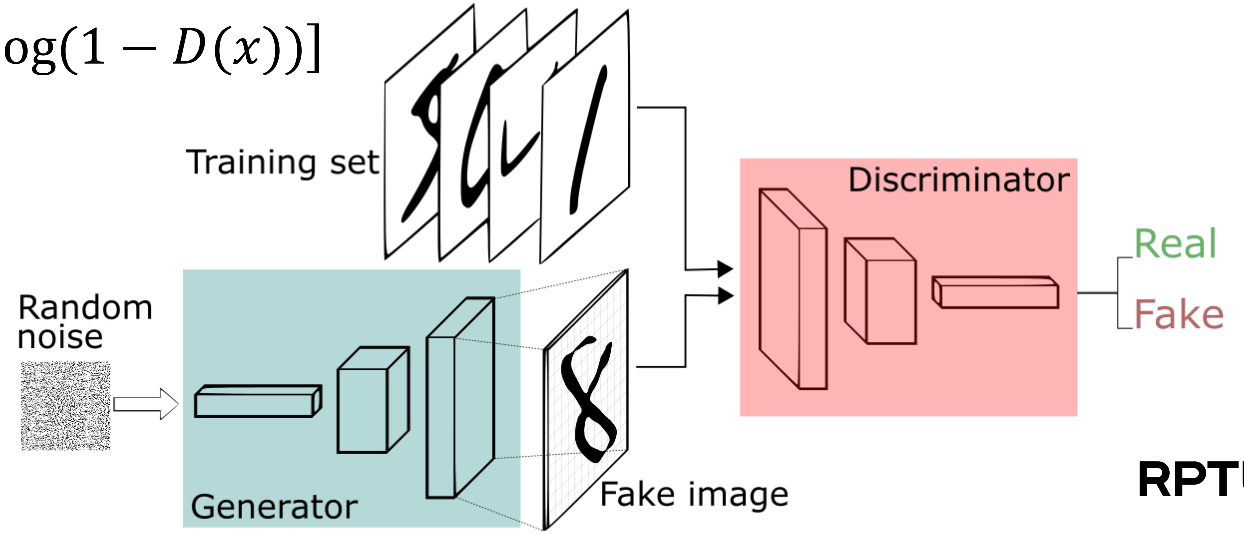
- [Goodfellow et al. 2014]
- Generative model $x = G_\theta(z), z \sim\ p(z)$
  - Map noise variable $z$ to data space $x$
- Discriminator $D_\phi(x)$
  - Output the probability that $x$ came from the data rather than the generator
- No explicit inference model
- No obvious connection to previous models with inference networks like VAEs

# Generative Adversarial Nets (GANs)

- Learning
  - A minimax game between the generator and the discriminator
  - Train $D$ to maximize the probability of assigning the correct label to both training examples and generated samples
  - Train $G$ to fool the discriminator

$$\max_D L_D = \max_D \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{x\sim G(z), z\sim p(z)}[\log(1 - D(x))]$$

$$\min_G L_G = \min_G \mathbb{E}_{x\sim G(z), z\sim p(z)}[\log(1 - D(x))]$$





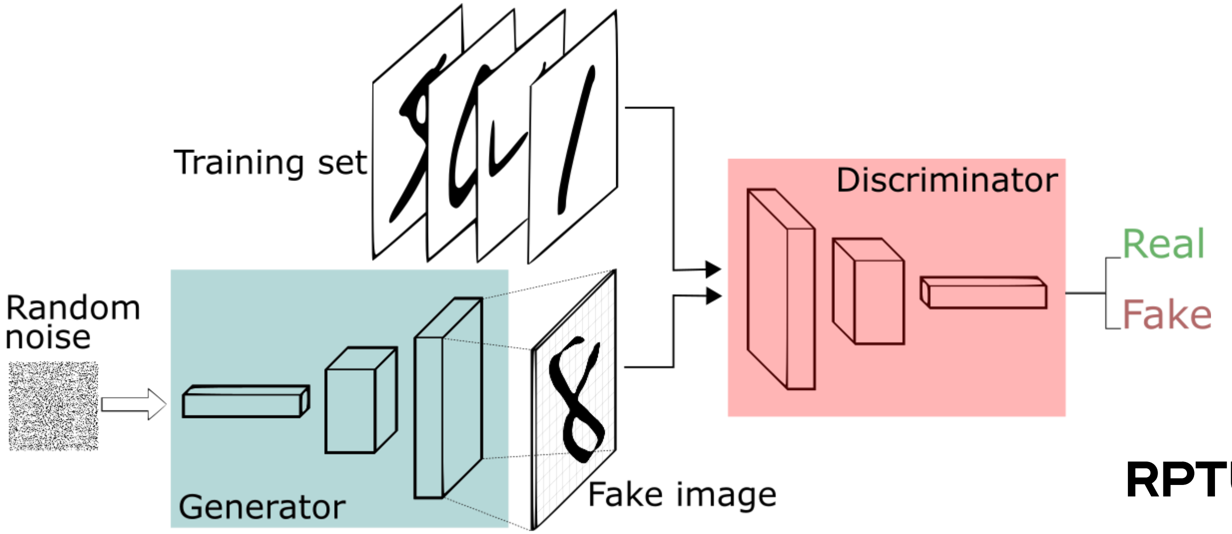

RPTU

# Generative Adversarial Nets (GANs)

$$\min_G L_G = \min_G \mathbb{E}_{x \sim G(z), z \sim p(z)}[\log(1 - D(x))]$$

- Learning
  - Train G to fool the discriminator
    - The original loss suffers from vanishing gradients when D is too strong
    - Instead use the following in practice

$$\max_G L_G = \mathbb{E}_{x \sim G(z), z \sim p(z)}[\log D(x)]$$

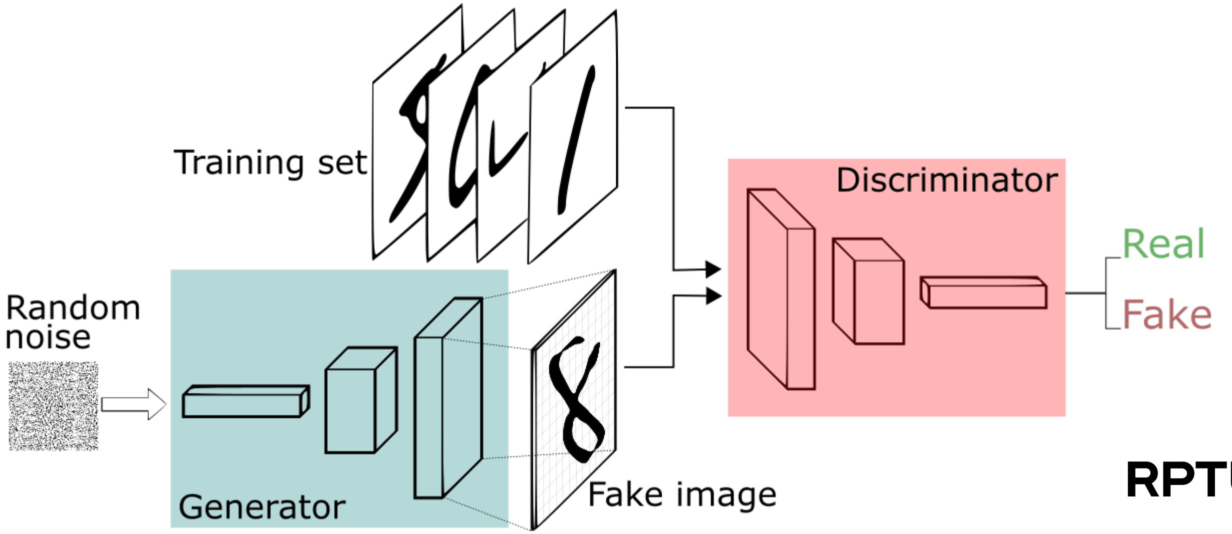





RPTU

# Generative Adversarial Nets (GANs)

- Learning
  - Aim to achieve equilibrium of the game
  - Optimal state:
    - $p_g(x) = p_{data}(x)$
    - $D(x) = \frac{p_{data}(x)}{p_{data}(x)+p_g(x)} = \frac{1}{2}$






RPTU

# Summary: GAN training

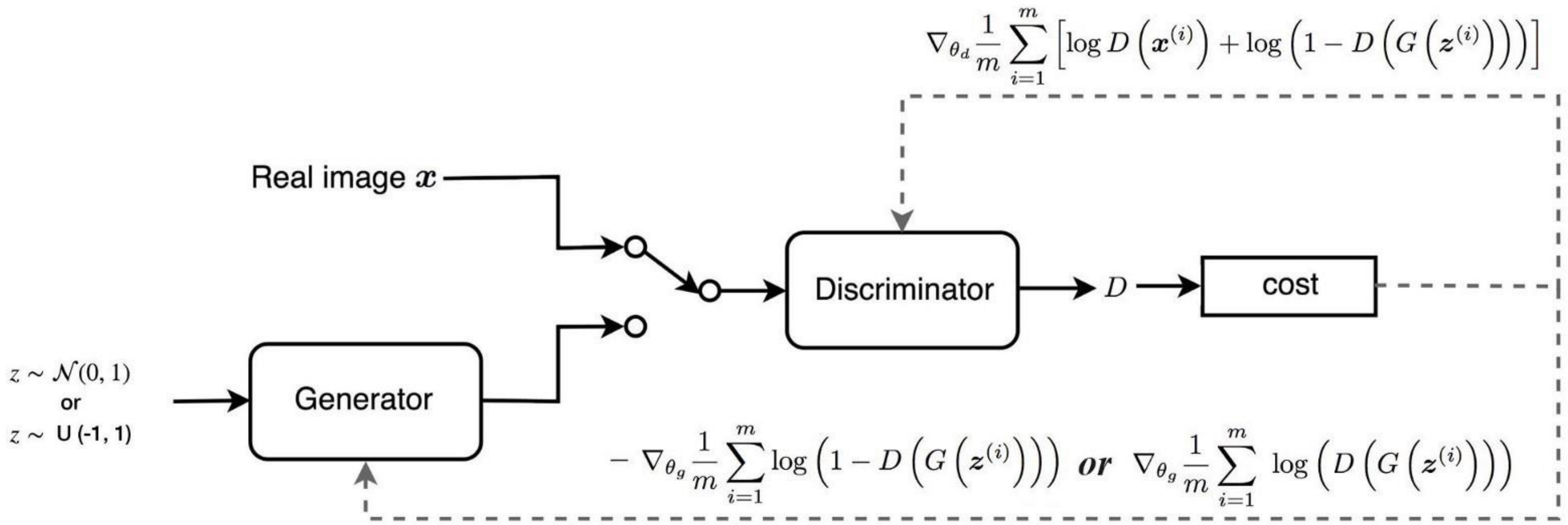
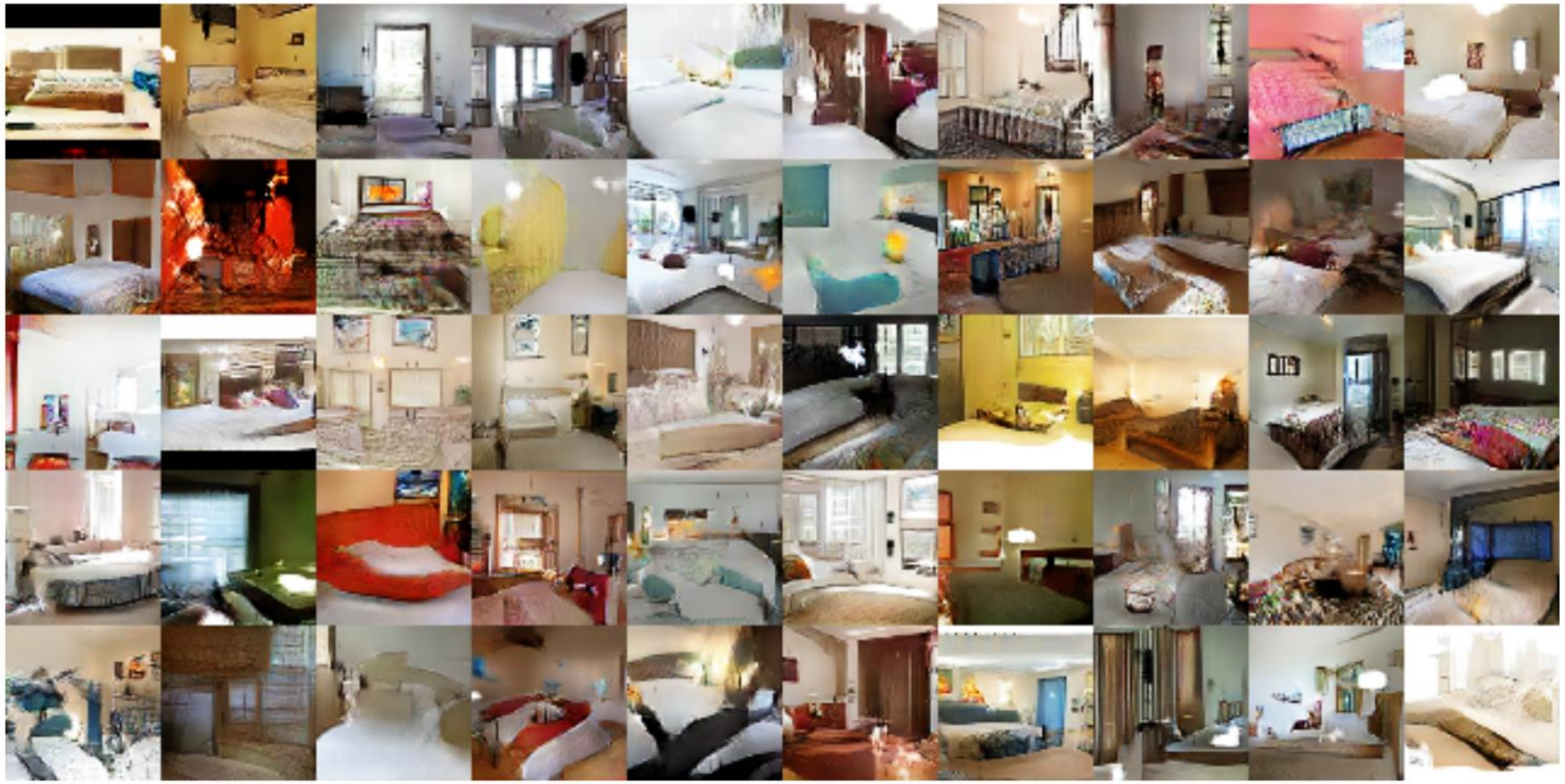


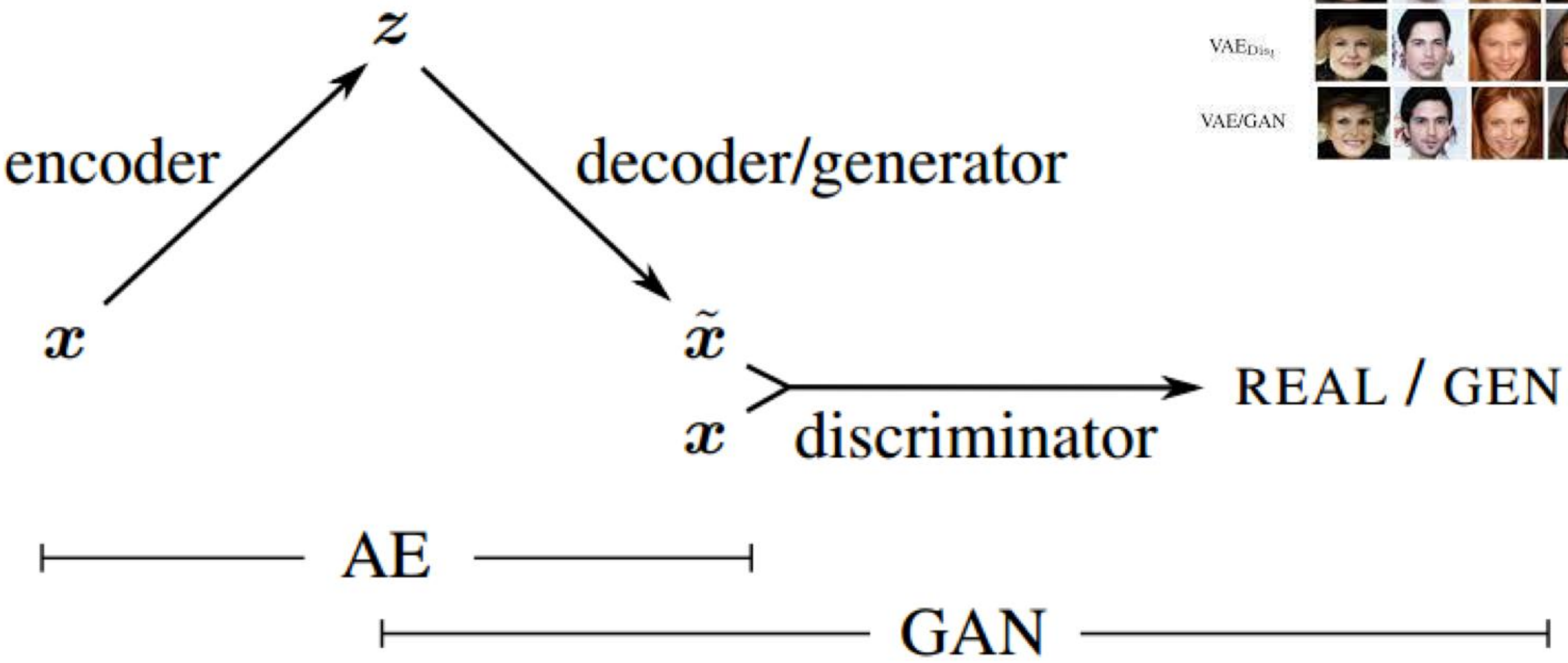

Image: Jonathan Hui

# GANs: Example Results



Generated bedrooms [Radford et al. 2016]

# VAE-GANs





Can potentially improve the blurriness of VAE outputs

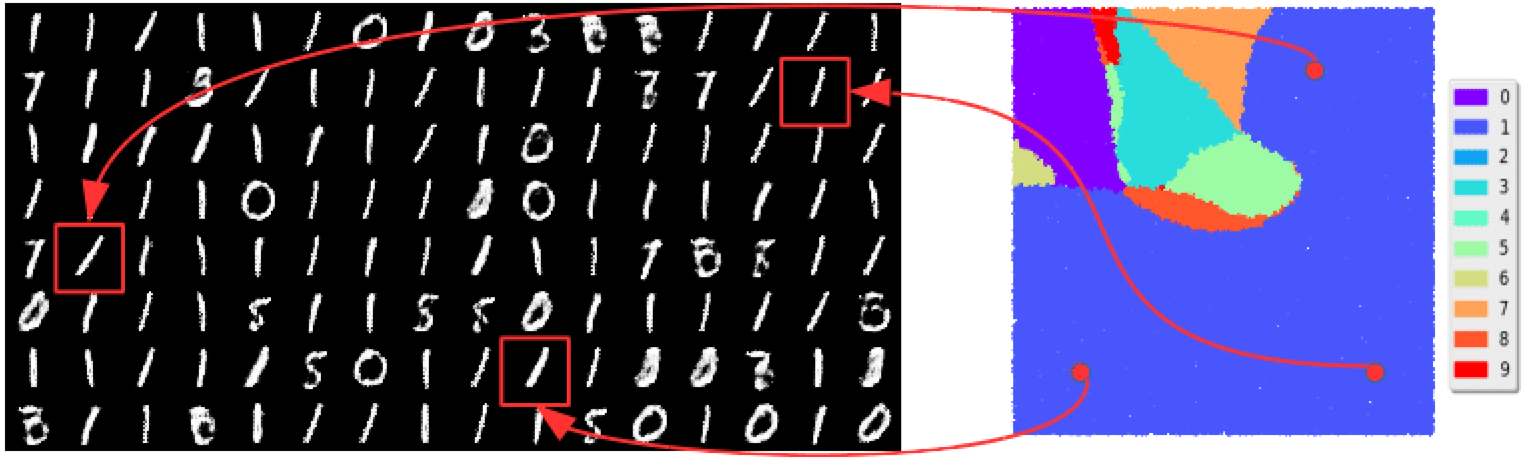


RPTU

[Larsen et al. 2015]

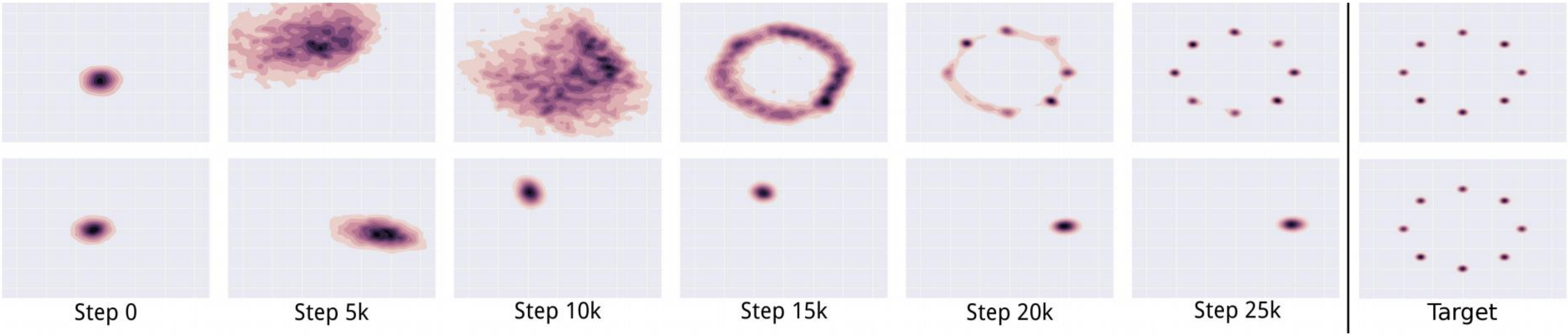
# Mode Collapse/Convergence issues

- Mode collapse refers to a phenomenon where only very similar images are generated.
  - If the discriminator does not change much, the best solution would be to continue to produce the one image that fools the discriminator most
- Optimization algorithms often approach a saddle point or local minimum rather than a global minimum
- Game solving algorithms may not approach an equilibrium at all




RPTU

# Mode Collapse




- The upper row shows a GAN that converges to the target distribution
- The lower row shows how the GAN only produces one mode and rotates to the next one as soon as the discriminator catches up
- Not always unwanted behavior: E.g. in style transfer we just need to find one good image rather than a diverse set of possible variants

# GAN Problems

- Non-convergence: model parameters oscillate, destabilize and never converge
- Mode collapse: the generator collapses which produces limited variety of examples
- Diminished gradient: The discriminator becomes too successful and the generator gradient vanishes and learns nothing
- Unbalance between generator and discriminator causing overfitting
- Highly sensitive to hyperparameter selection

# Text Style Transfer Introduction

Rheinland-Pfälzische
Technische Universität
Kaiserslautern
Landau

# Style is important

Image Source:
https://www.apple.com/de/siri/

Will it rain tomorrow?

Ain’t gonna rain, bro. Ya won’t need ya brolly.

It will not rain tomorrow, you will not need your umbrella, sir.

# Definition of text style

- Data-driven
  - Definition existing datasets used by the community
  - E.g. Amazon or Yelp reviews for sentiment transfer
- Linguistic
  - High level: Formality, simplification,…
  - Low level: Lexical, syntactic,…

# Examples for style transfer

You have to consider both sides of the story.
Gotta see both sides of the story.

**Formality [Rao et al. 2018]**

At the first God made the heaven and the earth.
In the beginning God created the heavens and the earth.

**Simplification [Carlson et al. 2018]**

This is just awful.
*This is pure genius.*

**Sentiment [Shen et al. 2017]**

# Style transfer models

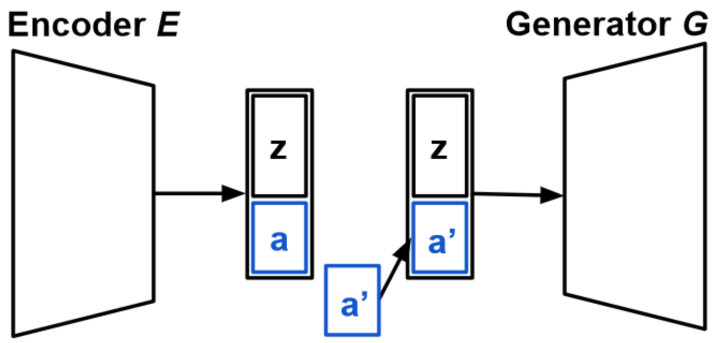
- Supervised models use style labels
  - Parallel methods
  - Non-parallel methods

- Unsupervised models do not use style labels

# Parallel text style transfer

- Usually, adopting seq2seq models from neural machine translation
  - Bi-directional LSTM + attention [Bahdanau et al. 2015] used by Rao et al. [2018] and Jhamtani et al. [2017]
  - Transformer-based [Vaswani et al. 2017]
- Data augmentation using back-translation to expand the dataset [Rao et al. 2018]

# Latent representation manipulation




- Latent representation splitting (e.g. John et al. [2019])
  - Disentangle latent representation into semantic representation $z$ and attribute (style) representation $a$
  - Replace $a$ by $a'$
  - Decode for style transfer

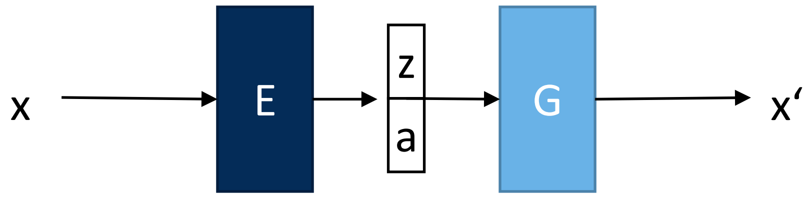


RPTU

# Training objectives

- Target attribute is fully and exclusively controlled by $a$
  - ➢Style-oriented losses

- Attribute-independent information is fully and exclusively controlled and captured by $z$
  - ➢Content-oriented losses






RPTU

# Style-oriented losses

- Attribute classifier on outputs: Make output carry target attribute $a'$ according to pre-trained classifier $f_c$ [Prabhumoye et al. 2018]

$$\mathcal{L}_{ACO}(\theta_G, a') = -\mathbb{E}_{p(x)} \log f_c(x')$$

- Attribute classifier on representations: Enforce style in hidden representation [John et al. 2019]

$$\mathcal{L}_{ACR}\left(\theta_E, \theta_{f_c}\right) = -\mathbb{E}_{p(x)} \log f_c(a)$$



RPTU

# Style-oriented losses

- Adversarial learning on representations: Enforce $z$ to not contain any information about $a$ [John et al. 2019]

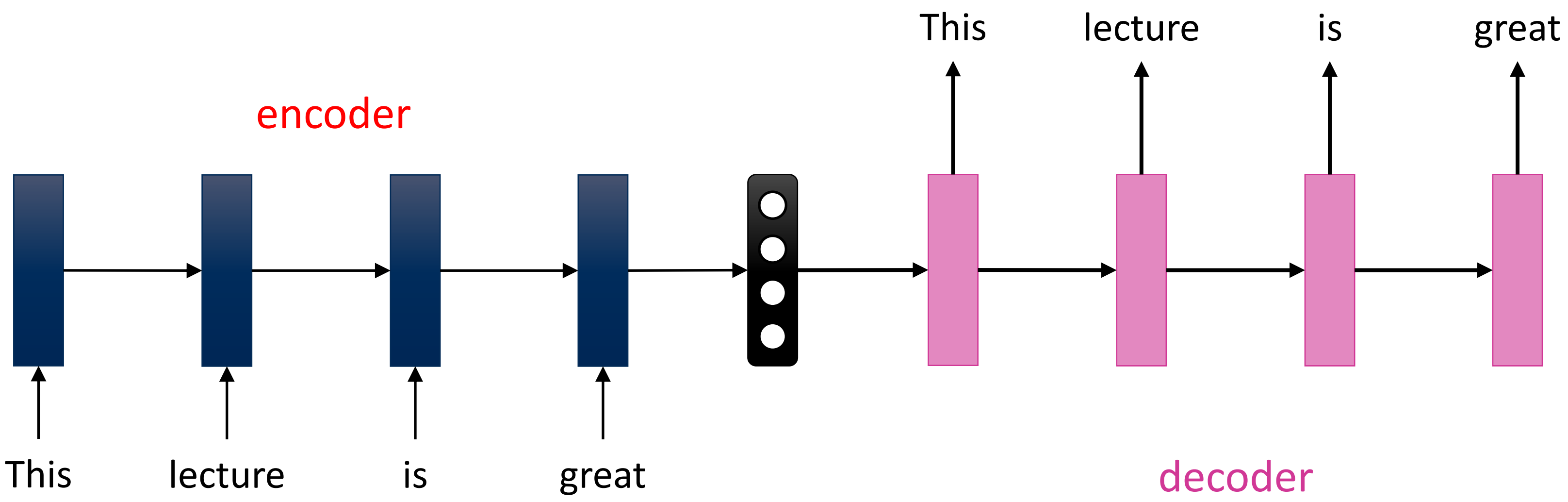
$$\max_{E} \min_{f_c} \mathcal{L}_{AdvR}\left(\theta_E, \theta_{f_c}\right) = -\mathbb{E}_{p(x)} \log f_c(E(x))$$



RPTU

# Educating Text Autoencoders: Latent Representation Guidance via Denoising

- Text autoencoders represent sentences as vectors in latent space

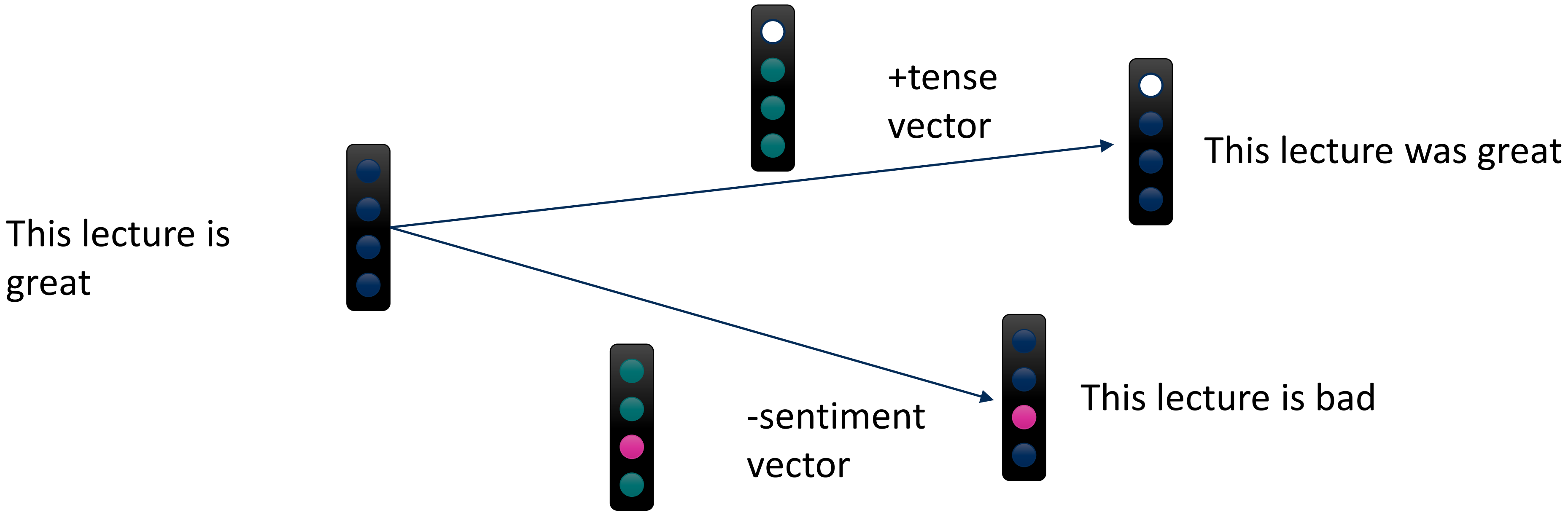



[Shen et al. 2020]



RPTU

# Manipulate sentences by modifying the latent representation



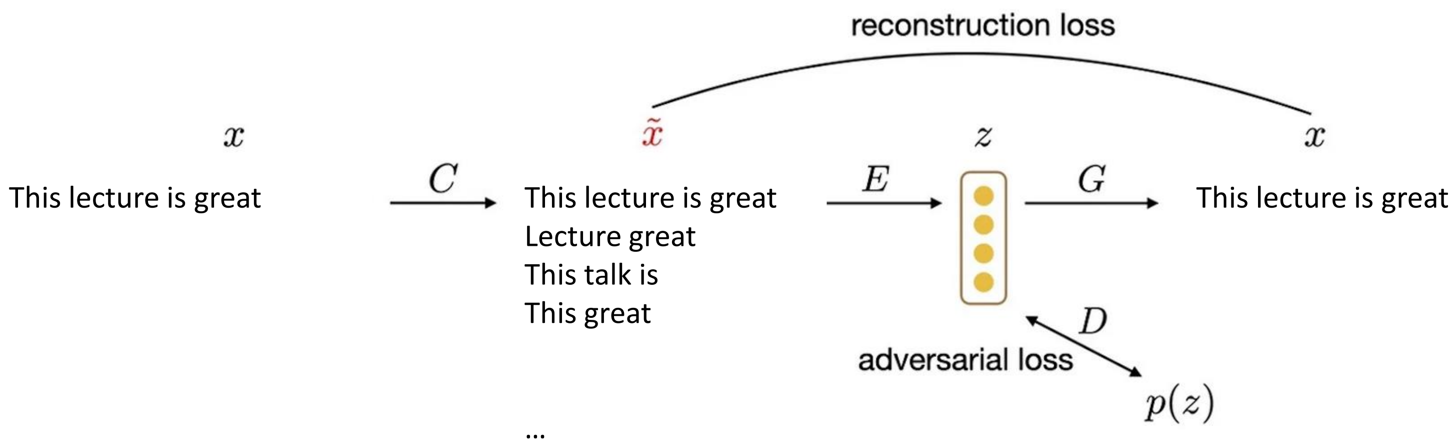

[Shen et al. 2020]



RPTU

# Denoising adversarial autoencoder

- Add perturbation process C that maps $x$ to nearby $\tilde{x}$ and ask model to reconstruct $x$ from $\tilde{x}$




$$\min_{E,G} \max_{D} \mathcal{L}_{rec}(\theta_E, \theta_G) - \lambda \mathcal{L}_{adv}(\theta_E, \theta_G)$$

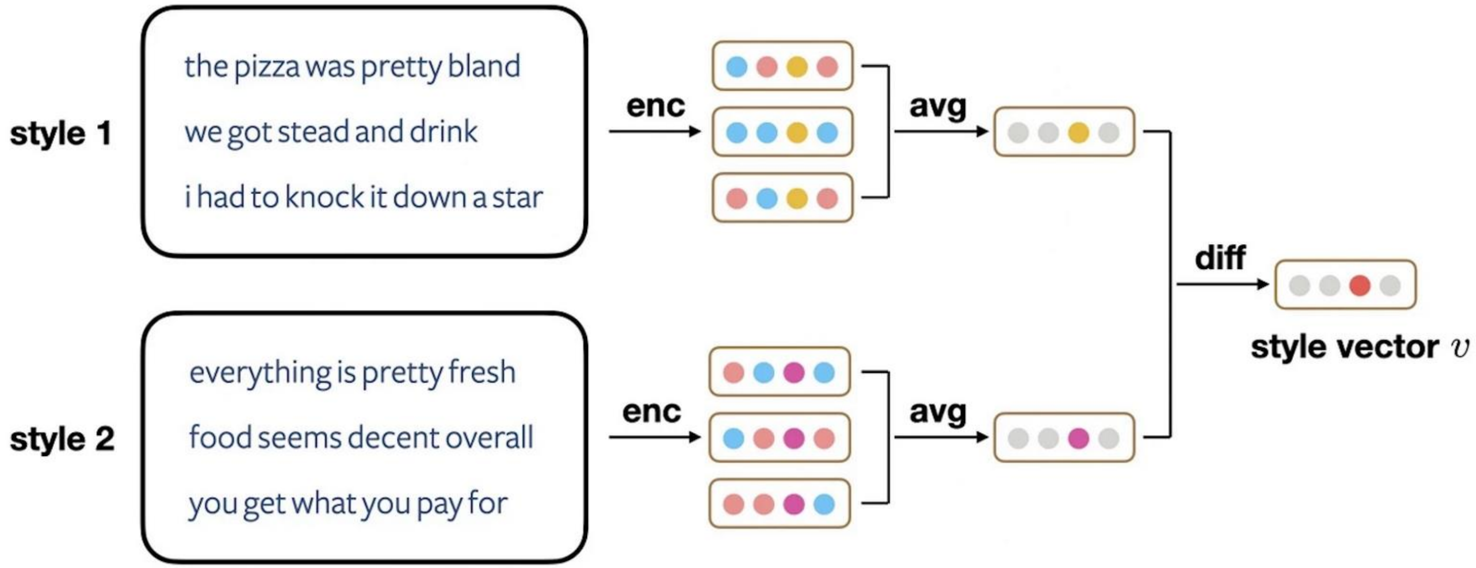
[Shen et al. 2020]

The discriminator D ensures that the latent encodings $z_1, \dots, z_n$ of training examples $x_1, \dots, x_n$ are indistinguishable from prior samples $z \sim p(z)$



RPTU

# Unsupervised style transfer with DAAE




[Shen et al. 2020]



RPTU

# Arbitrary text style transfer with large language models

| Original | I decided to say hello to him, and we stood there until he had to go. |
|---|---|
| more melodramatic | I decided to go meet him in the pouring rain, to declare my undying love for him in the rain. |
| more comic | I decided to make a funny face at the creepy man who was starring and we stood there making faces at each other until he had to go. |
| include the word "balloon" | I decided that he needed to be cheered up as well, and we stood there and talked until he had all the smiles he needed. Then he gave me a balloon, before he left. |
| include the word "park" | I decided to go for a walk in the park and there I met a man who had an uncanny resemblance to my friend. After that serendipitous encounter we went our seperate ways. |
| includes a metaphor | A snowflake landed on his nose and melted, and that was my cue to leave. |

[Reif et al. 2022]

# Arbitrary text style transfer with large language models

- Idea: Use natural language to describe the task (text style transfer)
- Prompts
  - Zero-Shot: transfer without examples
  - Few-Shot: transfer with few examples of the wanted style transferred
  - Augmented Zero-Shot: transfer with few examples of arbitrary styles

[Reif et al. 2022]

# Arbitrary text style transfer with large language models

- Zero-Shot
  - Here is some text: {That is an ugly dress}. Here is a rewrite of the text, which is more positive: {
- Few-Shot
  - **Here** is some text: {I was really sad about the loss}. Here is a rewrite of the text, which is more positive: {I was able to accept and work through the loss to move on.} **Here** is some text: {The eggnog was tasteless}. Here is a rewrite of the text, which is more positive: {The eggnog had a great, festive taste to it.} **...** **Here** is some text: {That is an ugly dress}. Here is a rewrite of the text, which is more positive: {

[Reif et al. 2022]

# Arbitrary text style transfer with large language models

- Augmented Zero-Shot
  - **Here** is some text: {When the doctor asked Linda to take the medicine, he smiled and gave her a lollipop}. Here is a rewrite of the text, which is more scary: {When the doctor told Linda to take the medicine, there had been a malicious gleam in her eye that Linda didn't like at all} **Here** is some text: {They asked loudly, over the sound of the train}. Here is a rewrite of the text, which is more intense: {They yelled aggressively, over the clanging of the train} **...** **Here** is some text: {That is an ugly dress}. Here is a rewrite of the text, which is more positive:{
- Augmented Zero-Shot works best in this setting
- Tested with several LLMs, LLM optimized for dialogs works best

[Reif et al. 2022]



RPTU

# Style transfer evaluation

- Dimensions
  - Fluency
  - Content Preservation
  - Style Transfer Accuracy
- Human annotation or automated metrics

# Conclusion

- Different definitions of text style
- Text style transfer and its evaluation easy on parallel data that is scarce
- Non-parallel methods
    - Disentanglement-based
    - Prototype editing
    - Pseudo-parallel corpus construction
- Evaluation conducted on three dimensions



RPTU

# References


- Rao, Sudha, and Joel Tetreault. "Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer." arXiv preprint arXiv:1803.06535 (2018).
- Carlson, Keith, Allen Riddell, and Daniel Rockmore. "Evaluating prose style transfer with the Bible." Royal Society open science 5.10 (2018): 171920.
- Shen, Tianxiao, et al. "Style transfer from non-parallel text by cross-alignment." Advances in neural information processing systems. 2017.
- DiMarco, Chrysanne, and Graeme Hirst. "A computational theory of goal-directed style in syntax." Computational Linguistics 19.3 (1993): 451-500.
- Lyu et al., “StylePTB: A Compositional Benchmark for Fine-Grained Controllable Text Style Transfer.” NAACL-HLT (2021).
- Jhamtani, Harsh, et al. "Shakespearizing modern language using copy-enriched sequence-to-sequence models." arXiv preprint arXiv:1707.01161 (2017).
- Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate. ICLR 2015
- Liu, Dayiheng, et al. "Revision in continuous space: Unsupervised text style transfer without adversarial learning." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 05. 2020.
- John, Vineet, et al. "Disentangled representation learning for non-parallel text style transfer." ACL (2019).
- Prabhumoye et al., “Style Transfer Through Back-Translation.” ACL (2018).
- Yang et al., “Unsupervised Text Style Transfer Using Language Models as Discriminators.” NeurIPS (2018)
- Logeswaran, Lee, and Bengio, “Content Preserving Text Generation with Attribute Controls.” NeurIPS (2018)




RPTU

# References


- Li, Juncen, et al. "Delete, retrieve, generate: a simple approach to sentiment and style transfer." NAACL-HLT (2018).
- Jin et al., “IMaT: Unsupervised Text Attribute Transfer via Iterative Matching and Translation.” EMNLP (2019)
- Hoang et al., “Iterative Back-Translation for Neural Machine Translation.”NMT@ACL 2018
- Shen et al., “Educating Text Autoencoders.” ICML (2020).
- Makhzani et al., “Adversarial Autoencoders.”
- Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- Mir et al., “Evaluating Style Transfer for Text.” NAACL-HLT (2019).
- Pang and Gimpel, “Unsupervised Evaluation Metrics and Learning Criteria for Non-Parallel Textual Transfer.” Proceedings of the 3rd Workshop on Neural Generation and Translation (2019)
- Zhang et al., “BERTScore: Evaluating Text Generation with BERT.” ICLR (2020).
- Kusner, Matt, et al. "From word embeddings to document distances." International conference on machine learning. PMLR, 2015.
- Kim, Yoon. “Convolutional Neural Networks for Sentence Classification.” EMNLP (2014).
- Reif, Emily, et al. "A recipe for arbitrary text style transfer with large language models." ACL (2022)
- Jin et al., “Deep Learning for Text Style Transfer: A Survey.”
- Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002 Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002.




RPTU

# References


- Bengio, Yoshua et al. “A Neural Probabilistic Language Model.” J. Mach. Learn. Res. (2000).
- Mikolov, Tomas et al. “Efficient Estimation of Word Representations in Vector Space.” ICLR (2013).
- Kiros, Ryan et al. “Skip-Thought Vectors.” NIPS (2015)
- Devlin, Jacob et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” NAACL (2019)
- Gutmann, Michael, and Aapo Hyvärinen. "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models." Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010.
- OORD, Aaron van den; LI, Yazhe; VINYALS, Oriol. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.
- Jaiswal, Ashish, et al. "A survey on contrastive self-supervised learning." Technologies 9.1 (2020): 2
- Logeswaran, Lajanugen, and Honglak Lee. "An efficient framework for learning sentence representations." ICLR (2018)
- Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International Conference on Machine Learning. PMLR, 2021.
- Liu, Xiao, et al. "Self-supervised learning: Generativ"A survey on contrastive self-supervised learning.e or contrastive." IEEE Transactions on Knowledge and Data Engineering (2021).
- Jaiswal, Ashish, et al. " Technologies 9.1 (2020): 2.
- Le-Khac, Phuc H., Graham Healy, and Alan F. Smeaton. "Contrastive representation learning: A framework and review." IEEE Access 8 (2020): 193907-193934.




RPTU