A Subtle Introduction to Deep Representation Learning

Applied Machine and Deep Learning
190.015

M.Sc. Fotios (Fotis) Lygerakis

October 2023

Chair of Cyber-Physical-Systems

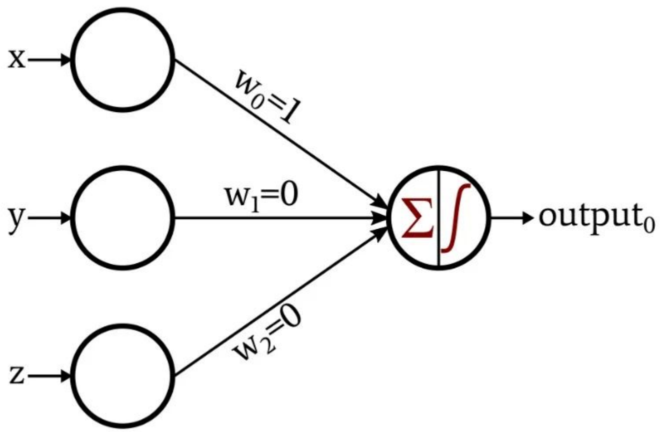
Outline

- Recap on Neural Networks
- Introduction to Representation Learning
- Core Techniques and Approaches
- Case Studies
- Coding Examples
- Conclusion and Q&A

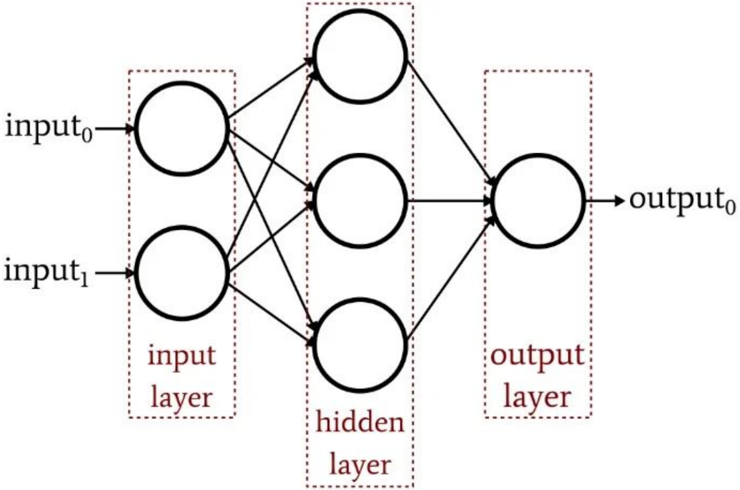
Recap on Neural Networks

Recap: Neuron & Neural Network

Perceptron



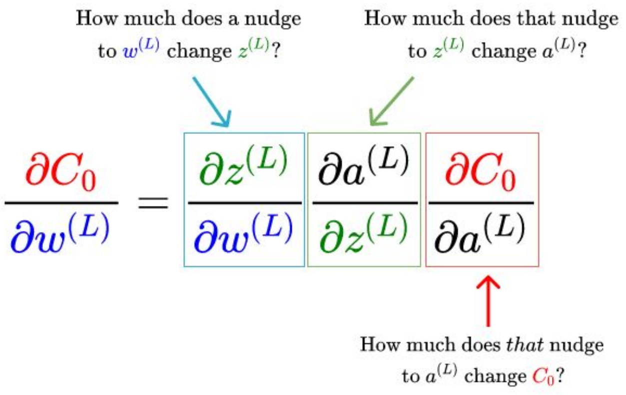
Deep Neural Network



https://www.allaboutcircuits.com/technical-articles/how-to-train-a-basic-perceptron-neural-network/

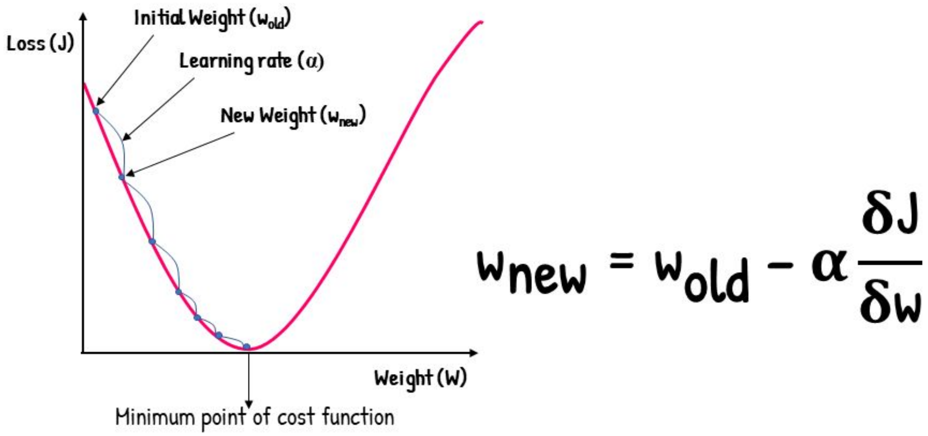
Recap: BackProp and Gradient Descent

Backpropagation



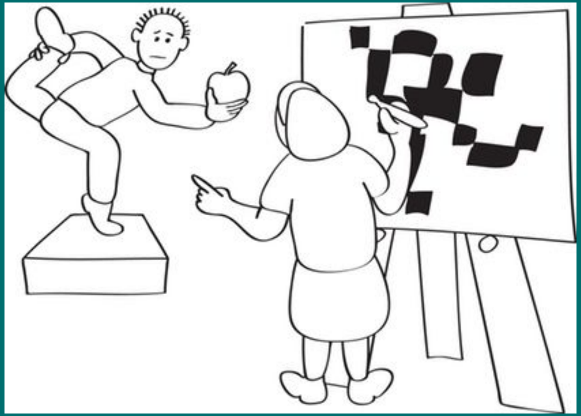
https://www.3blue1brown.com/lessons/backpropagation-calculus

Gradient Descent



https://www.analyticsvidhya.com/blog/2023/01/gradient-descent-vs-backpropagation-whats-the-difference/

Representation Learning



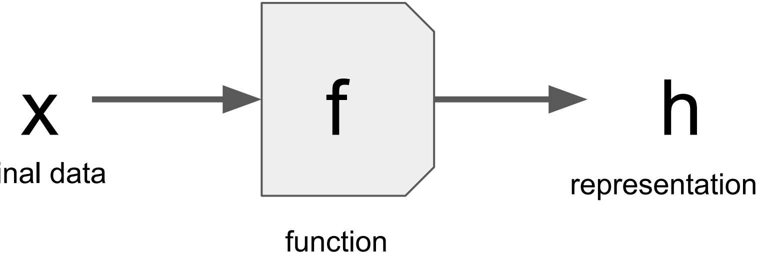
https://classic.csunplugged.org/activities/image-representation/

Definition

“Representation Learning is a process in machine learning where algorithms extract meaningful patterns from raw data to create representations that are easier to understand and process. These representations can be designed for interpretability, reveal hidden features, or be used for transfer learning.”

https://paperswithcode.com/task/representation-learning

$$h = f(x)$$



Reads

- Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. **Representation Learning: A Review and New Perspectives**. IEEE Trans. Pattern Anal. Mach. Intell. 35, 8 (August 2013), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Feature Engineering vs Representation Learning

GO* Feature Engineering

- Manually defined features
- Requires domain expertise
- time-consuming/not scalable with more data

Representation Learning

- Automatically learns the most important features
- No need for explicit programming
- More efficient and can handle complex, high-dimensional data

*GO: good old

Why is it important?

- **Highlight Essential Features:** Focuses the model's attention on the most important aspects of the data
- **Dimensionality Reduction**: Speeds up learning and helps removing noise, enhancing performance
- **Computational Efficiency:** Train faster
- **Improved Generalization:** Facilitates better understanding of unseen data, building robust models

Today’s Focus on **Representation Learning**

- Fundamentals
- Core Techniques & Approaches
- Applications in various domains

WHY Deep?

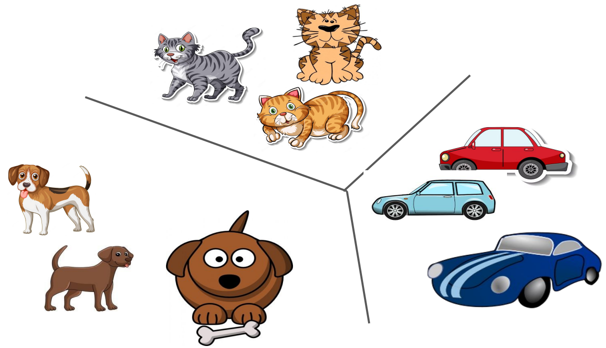
Automated feature engineering

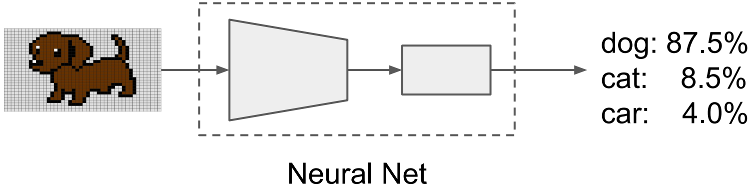
Improved Performance

Real-World Applications

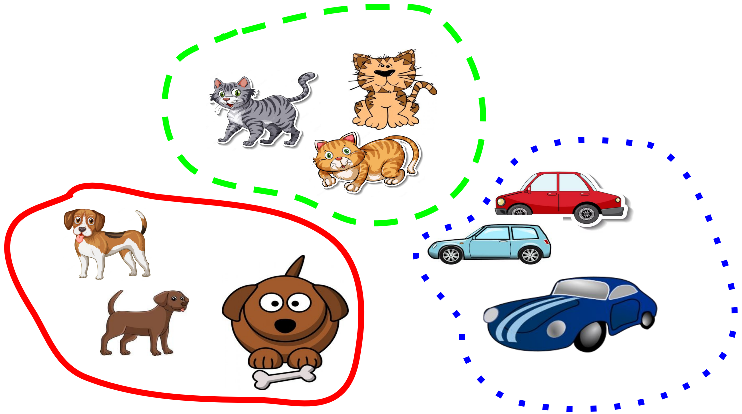
Kinds of Representations Learning

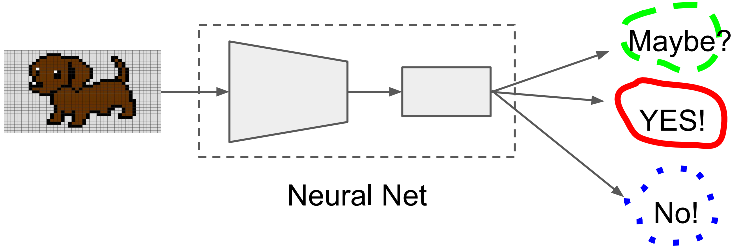
Supervised





Unsupervised

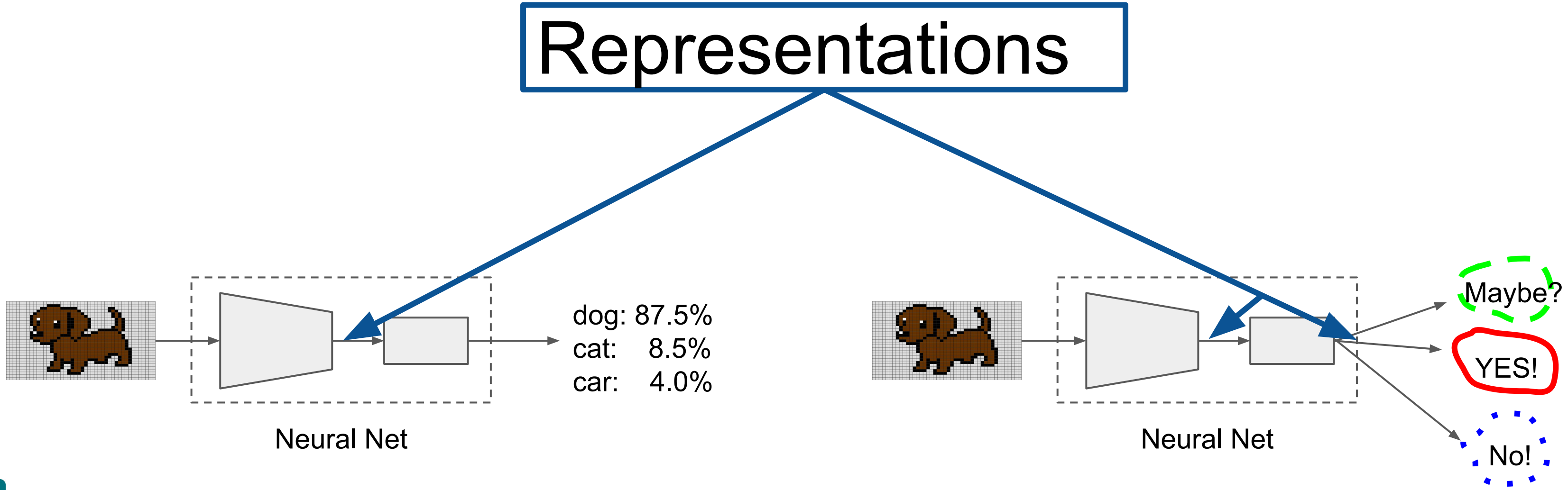




Kinds of Representations Learning

Supervised

Unsupervised



Which flavor to choose?

Supervised

- Benefits:
  - Learning from labeled data.
  - Often provides better performance with adequate labeled data.
- Best for:
  - Tasks with a substantial amount of labeled data.
  - Situations where high accuracy is essential and labeled examples are clear.

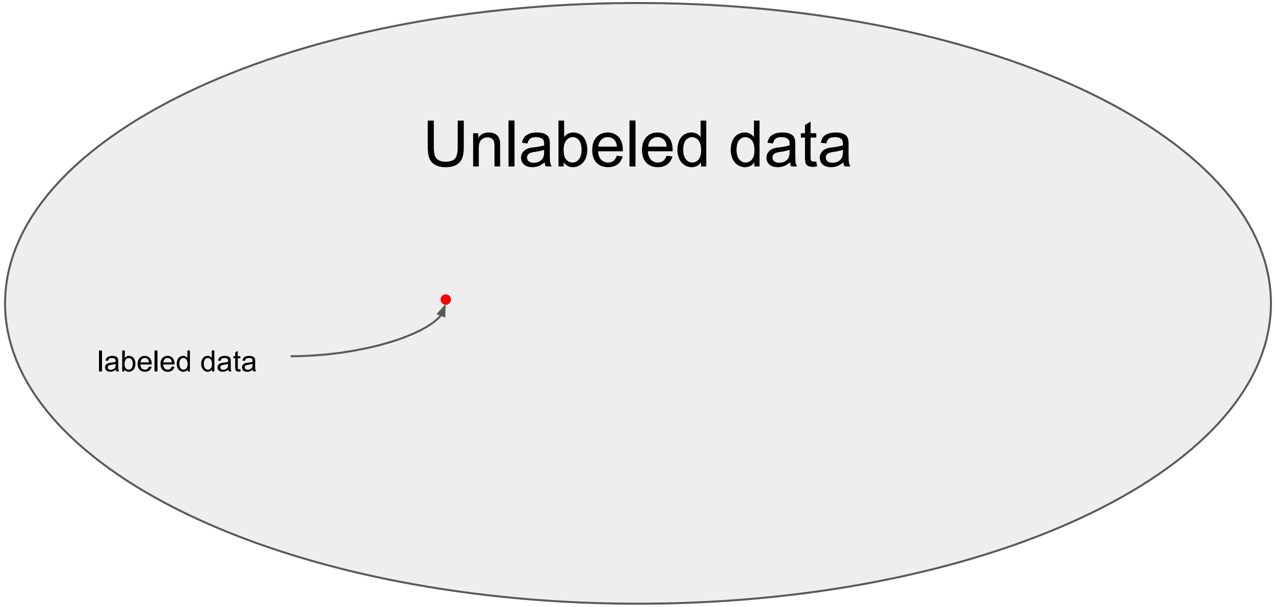
Unsupervised

- Benefits:
  - Learning from unlabeled data, **which is more abundant.**
  - Can uncover latent structures and features not evident from labels.
- Best for:
  - Tasks with limited or no labeled data.
  - Understanding the underlying structure of the data.

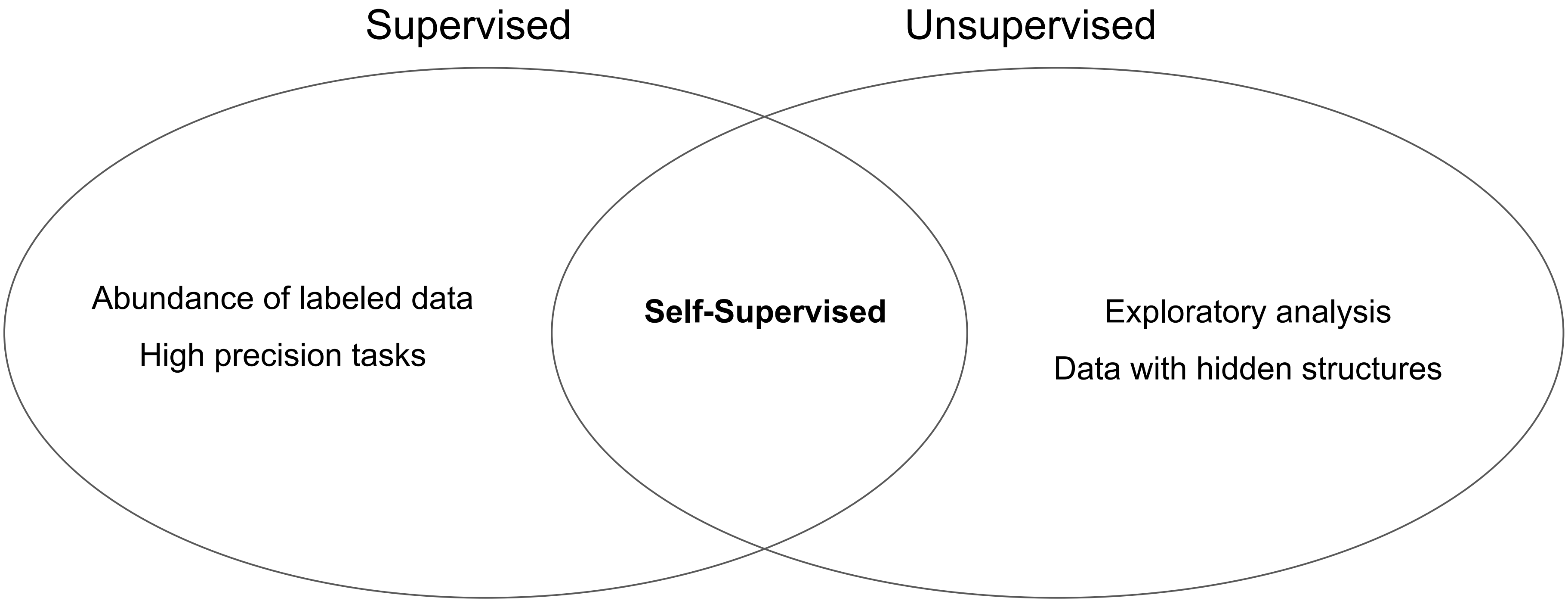
Which flavor to choose?

Supervised

Unsupervised



Which flavor to choose?



Self-Supervised Learning

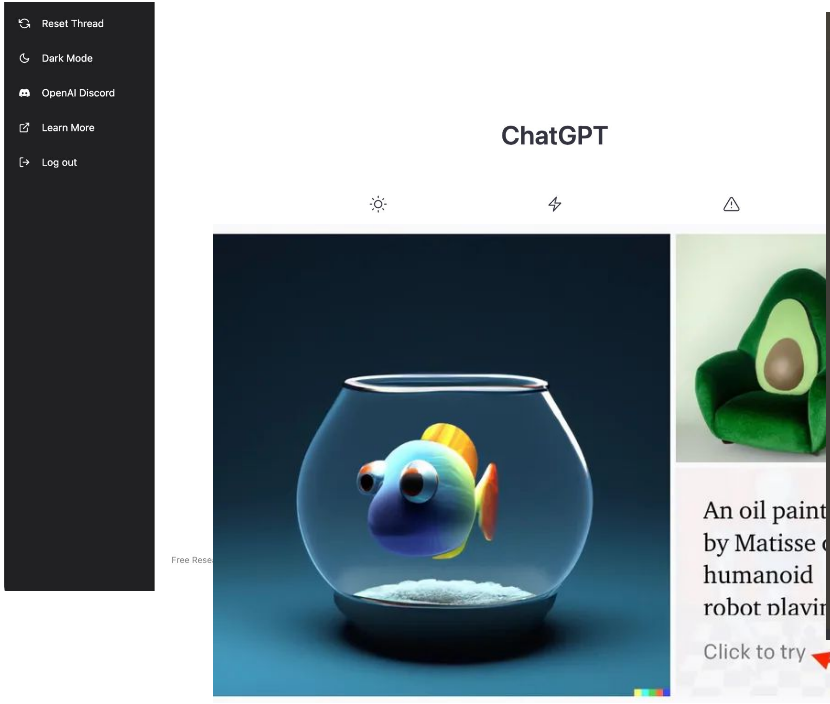
- Supervision is derived from the input data itself
- No explicit external labels needed
- Training by designing auxiliary tasks
- Given a portion of the data the model is trained to **predict** or **reconstruct** the remaining part.
- Once the model is trained in this manner, the learned representations can be used for downstream tasks, often with a separate fine-tuning step.

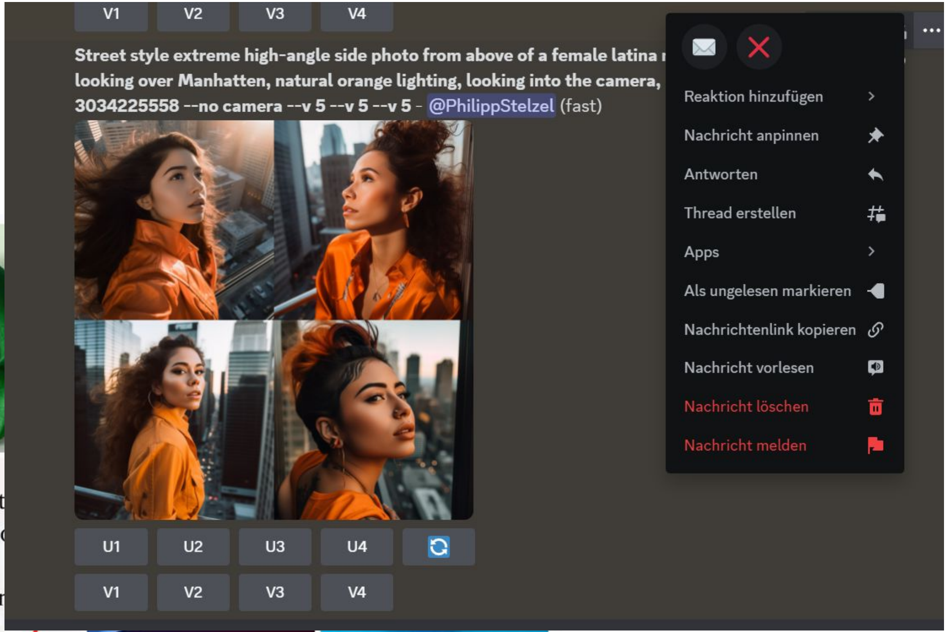
That's what we are going to discover today!

Real World Applications!

- Image Denoising
- Anomaly Detection
- Dimensionality Reduction
- Data Compression
- Visual Representation Learning
- Pre-training for Downstream Tasks
- Image Retrieval
- Image Generation
- Data Augmentation
- Super-Resolution
- Style Transfer
- Question Answering
- Multimodal Learning
- Robot Learning

Real World Applications!



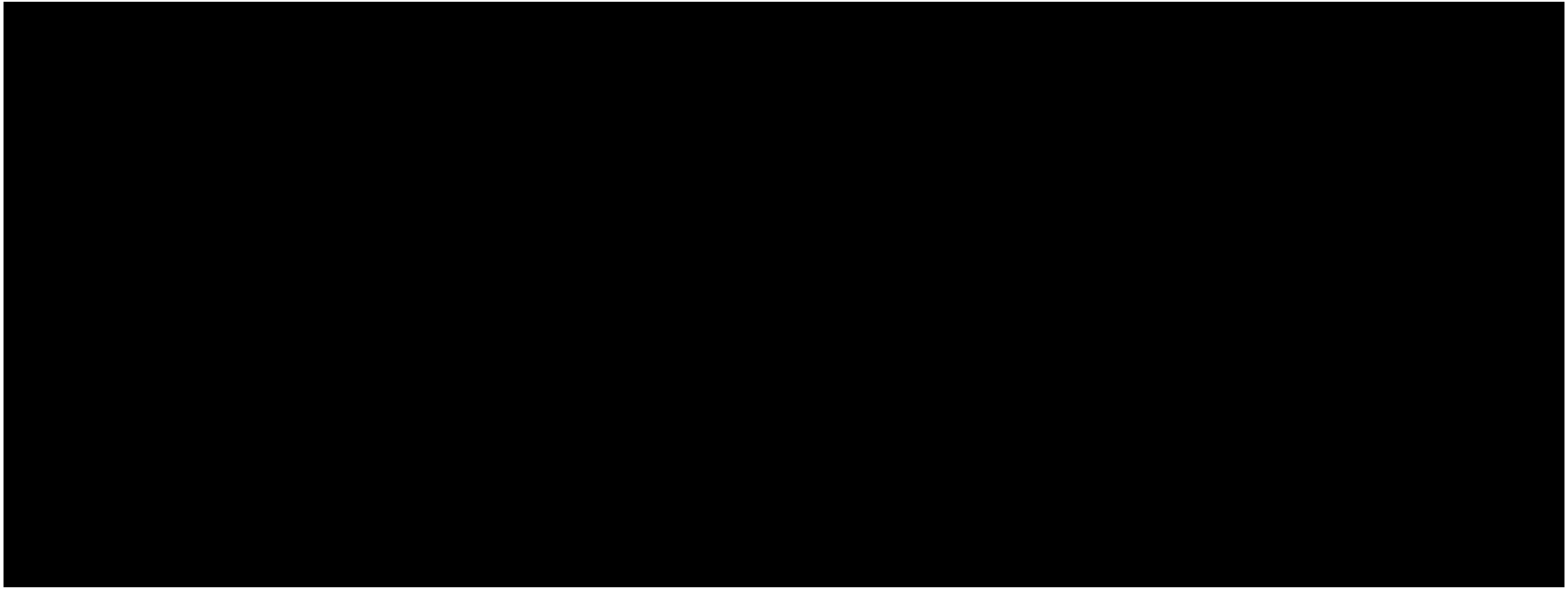


https://openai.com/chatgpt
https://openai.com/dall-e-2

https://www.midjourney.com
DINOv2

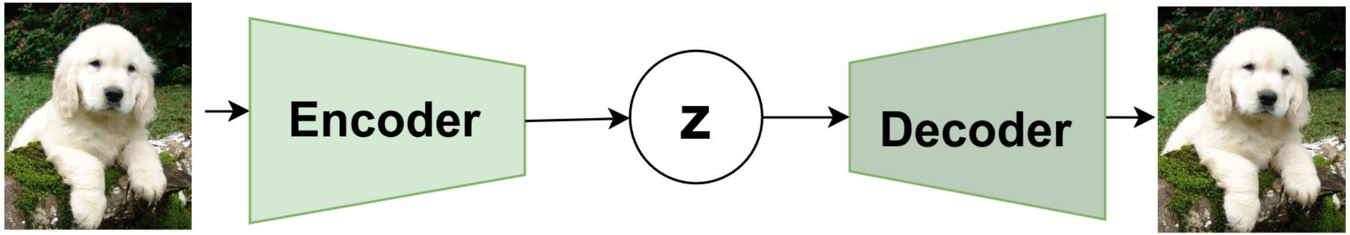
RT2

Real World Applications!



Core Techniques and Approaches

Autoencoders



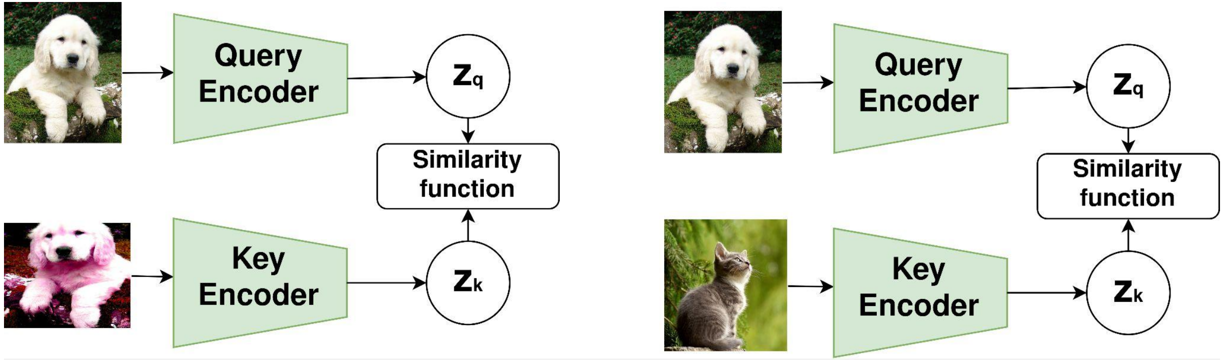
$$img_{original} \longrightarrow f_{encoder}(img) \longrightarrow z \longrightarrow f_{decoder}(z) \longrightarrow img_{reconstracted}$$

Training

1. Forward pass (left-to-right)
2. Mean Square Error($img_{original}$, $img_{reconstructed}$)
3. Backpropagate the MSE error

Reads
https://www.v7labs.com/blog/autoencoders-guide
https://www.geeksforgeeks.org/implementing-an-autoencoder-in-pytorch/
https://www.tutorialspoint.com/how-to-implementing-an-autoencoder-in-pytorch

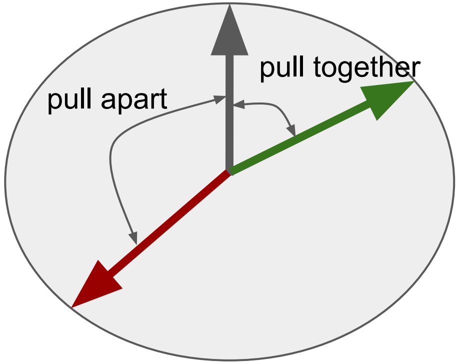
Contrastive Learning



$\text{img}_{\text{anqor}} \longrightarrow f_{\text{query}}(\text{img}) \longrightarrow z_q$

$\text{img}_{\text{positive}} \longrightarrow f_{\text{key}}(\text{img}) \longrightarrow z_{\text{k\_positive}}$

$\text{img}_{\text{negative}} \longrightarrow f_{\text{key}}(\text{img}) \longrightarrow z_{\text{k\_negative}}$



Reads
https://www.v7labs.com/blog/contrastive-learning-guide
https://encord.com/blog/guide-to-contrastive-learning/
https://www.youtube.com/watch?v=sftIkJ8MYL4

Contrastive Learning

Training

- Data Augmentation:
- Forward Pass:
  - Compute z_q
  - Compute z_k
- Compute Similarity
- Contrastive Loss L_{NCE}
- Backward Pass

$$\mathcal{L}_{NCE} = -\mathbf{E}\left[log \frac{e^{sim(z_i, z_j)}}{\sum_{k=1}^{K} e^{sim(z_i, z_k)}} \right]$$

Reads
https://arxiv.org/abs/2002.05709
https://arxiv.org/abs/1911.05722
https://arxiv.org/abs/2006.07733
https://arxiv.org/abs/2006.09882
https://arxiv.org/abs/2105.04906

Transformers

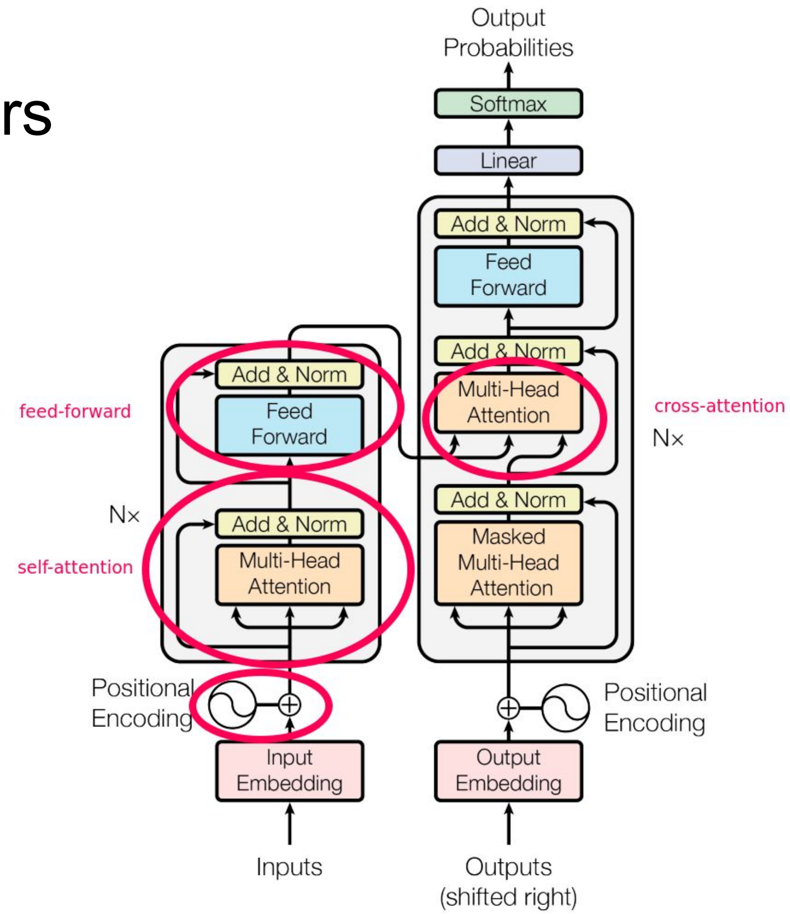
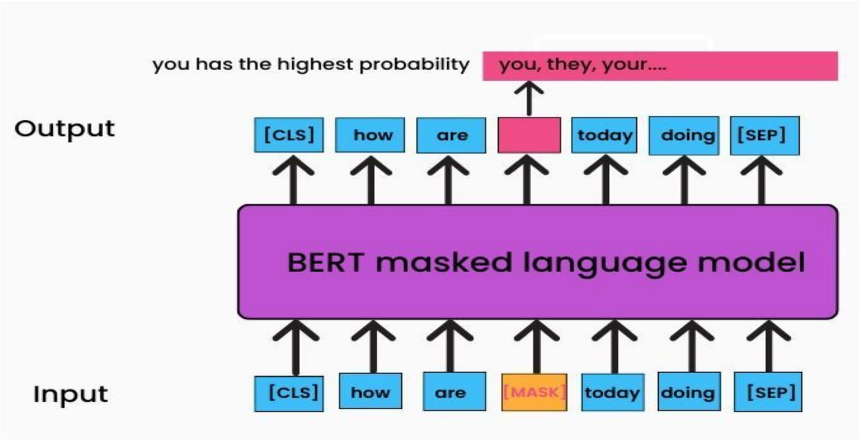
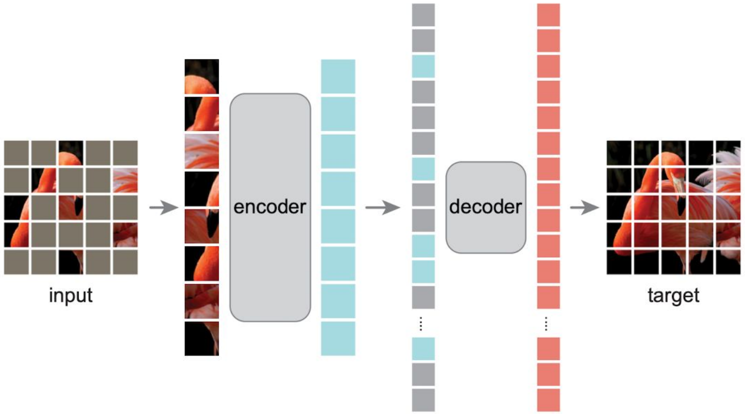


Figure 1: The Transformer - model architecture.

Masked Autoencoders

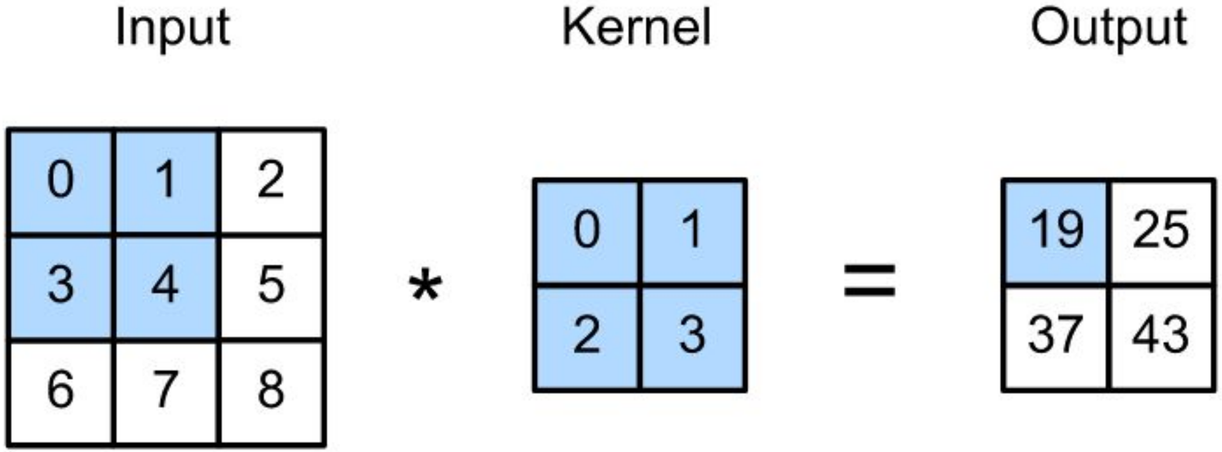




Reads
https://medium.com/dair-ai/papers-explained-28-masked-autoencoder-38cb0dbed4af
https://towardsdatascience.com/into-the-transformer-5ad892e0cee
https://towardsdatascience.com/illustrated-guide-to-transformers-step-by-step-explanation-f74876522bc0

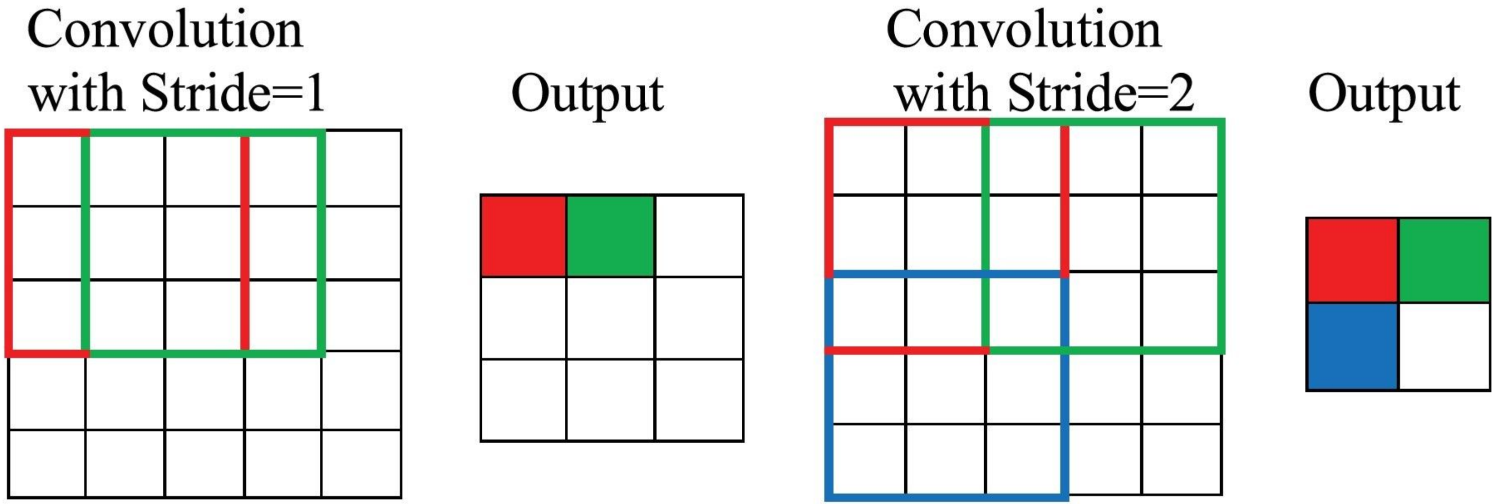
Basics of Convolution

Convolutions Basics: Kernel(filter)



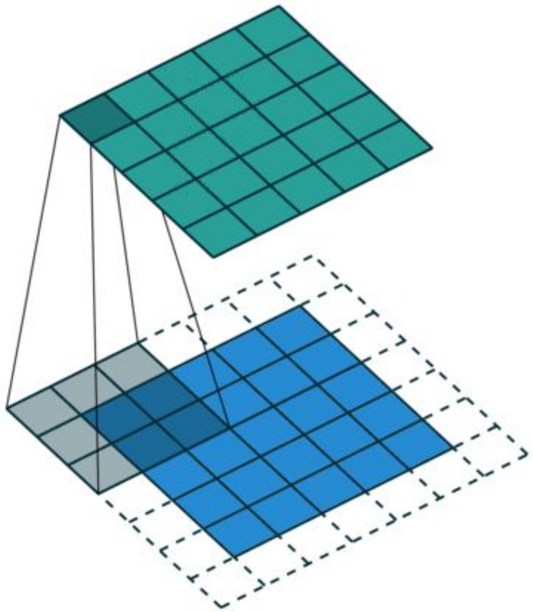
https://courses.cs.washington.edu/courses/cse446/21au/sections/08/convolutional_networks.html

Convolutions Basics: Stride



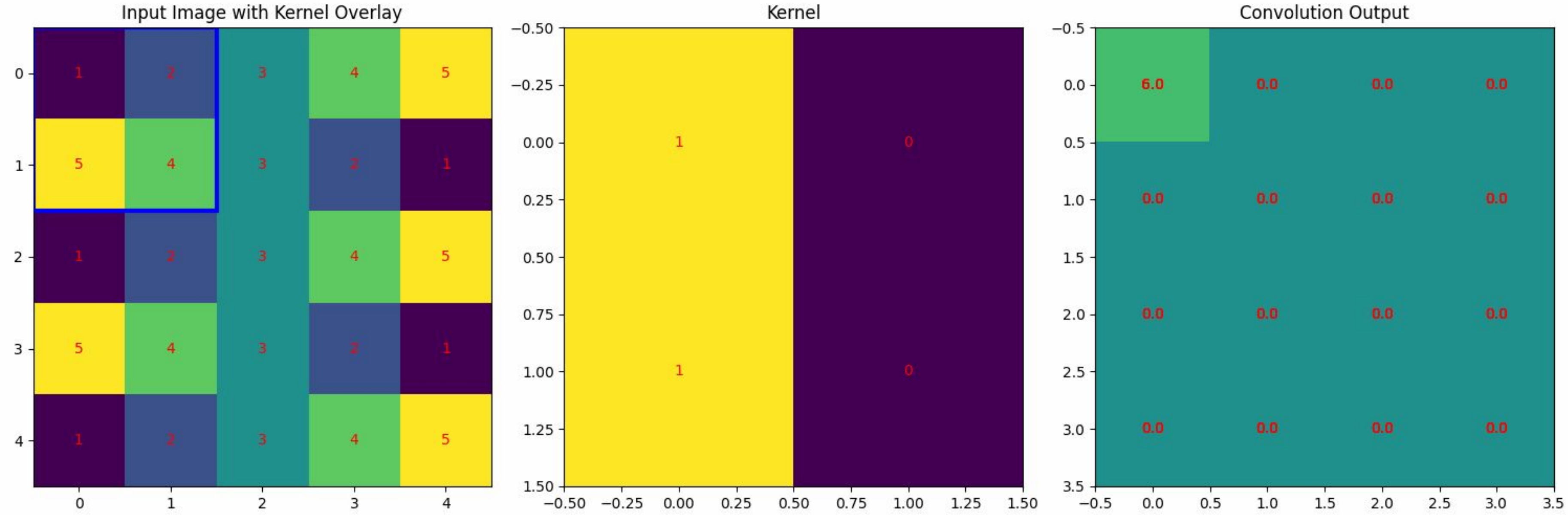
https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/

Convolutions Basics: Padding



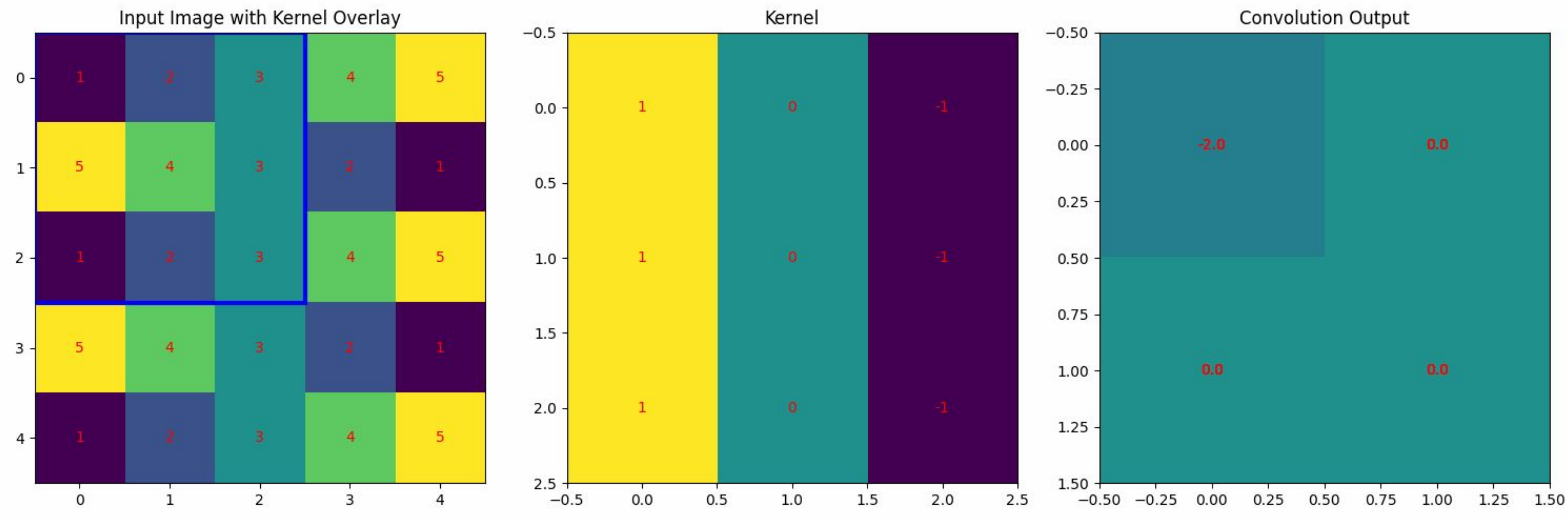
https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d

Examples



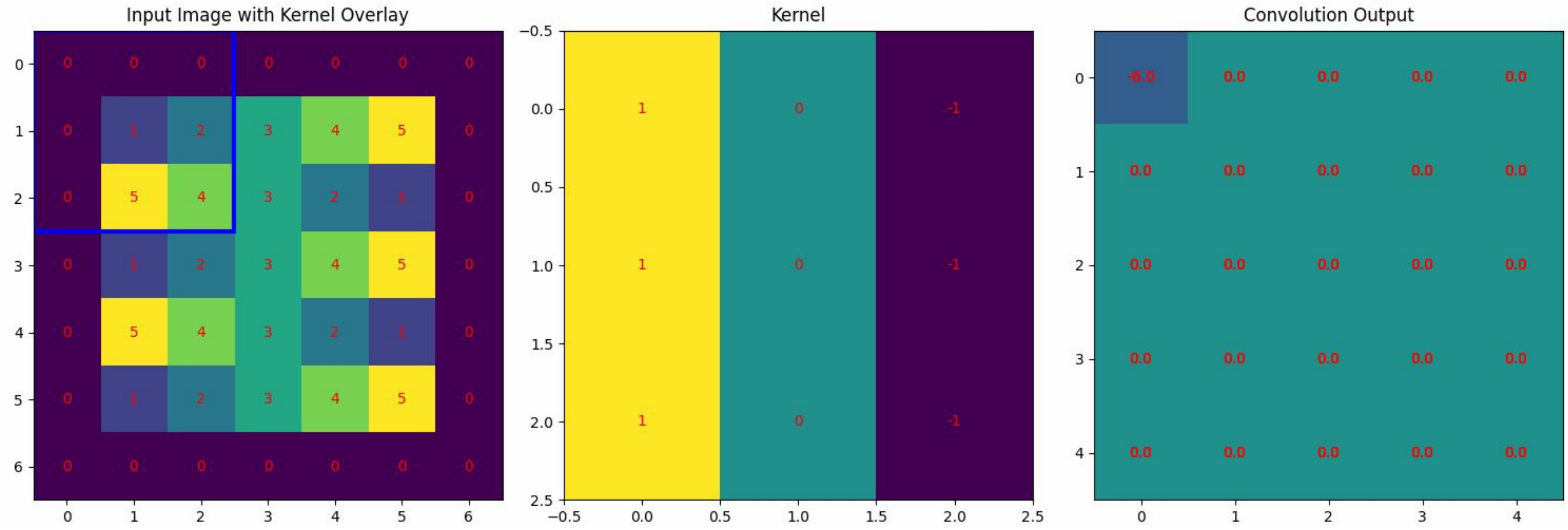
Kernel: 2x2 Stride: 1 Padding: 0

Examples



Kernel: 3x3 Stride: 2 Padding: 0

Examples



Kernel: 3x3 Stride: 1 Padding: 1

Coding Examples

Github Repository

Case Studies

Case Studies

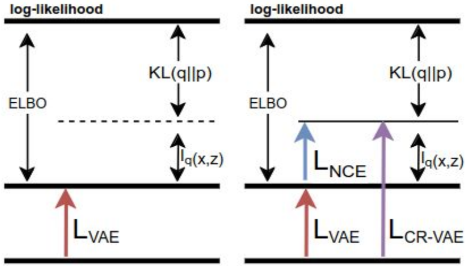
CR-VAE: Contrastive Regularization on Variational Autoencoders for Preventing Posterior Collapse

Fotios Lygerakis
Chair of Cyber-Physical Systems
University of Leoben
Austria
fotios.lygerakis@unileoben.ac.at
https://orcid.org/0000-0001-8044-3511

Elmar Rueckert
Chair of Cyber-Physical Systems
University of Leoben
Austria
elmar.rueckert@unileoben.ac.at
https://orcid.org/0000-0003-1221-8253

***Abstract*—The Variational Autoencoder (VAE) is known to suffer from the phenomenon of *posterior collapse*, where the latent representations generated by the model become independent of the inputs. This leads to degenerated representations of the input, which is attributed to the limitations of the VAE's objective function. In this work, we propose a novel solution to this issue, the *Contrastive Regularization for Variational Autoencoders (CR-VAE)*. The core of our approach is to augment the original VAE with a contrastive objective that maximizes the mutual information between the representations of similar visual inputs. This strategy ensures that the information flow between the input and its latent representation is maximized, effectively avoiding *posterior collapse*. We evaluate our method on a series of visual datasets and demonstrate, that CR-VAE outperforms state-of-the-art approaches in preventing *posterior collapse*. Code for this project is available at https://github.com/ligerfotis/crvae.**

***Index Terms*—variational autoencoders, contrastive learning, posterior collapse**



(a) Original ELBO (b) Contrastive Regularization

Fotios Lygerakis, & Elmar Rueckert. (2023). CR-VAE: Contrastive Regularization on Variational Autoencoders for Preventing Posterior Collapse. Accepted at ACAIT 2023

Wrapping Up

Wrap up

- Representation Learning is important
  - Highlight Essential Features
  - Dimensionality Reduction
  - Computational Efficiency
  - Improved Generalization
  - Exploit big unlabeled data
- Real-World Applications
- Core Techniques
  - Autoencoders
  - Contrastive Learning
  - Transformers

Resources & Extra Reads

Introduction to Pytorch

Autoencoders

https://www.v7labs.com/blog/autoencoders-guide

https://www.geeksforgeeks.org/implementing-an-autoencoder-in-pytorch/

https://www.tutorialspoint.com/how-to-implementing-an-autoencoder-in-pytorch

Contrastive Learning

https://lilianweng.github.io/posts/2021-05-31-contrastive/
https://arxiv.org/abs/2002.05709
https://arxiv.org/abs/1911.05722
https://arxiv.org/abs/2006.07733
https://arxiv.org/abs/2006.09882
https://arxiv.org/abs/2105.04906

Masked Autoencoders(Transformers)

https://medium.com/dair-ai/papers-explained-28-masked-autoencoder-38cb0dbed4af
https://towardsdatascience.com/into-the-transformer-5ad892e0cee
https://towardsdatascience.com/illustrated-guide-to-transformers-step-by-step-explanation-f74876522bc0

Thank you for your attention!

Fotios (Fotis) Lygerakis

Chair of Cyber-Physical-Systems

Montanuniversität Leoben

Franz-Josef-Straße 18,

8700 Leoben, Austria

E-mail: fotios.lygerakis@unileoben.ac.at

Web: https://cps.unileoben.ac.at/fotios-lygerakis-m-sc/

Presentation Link