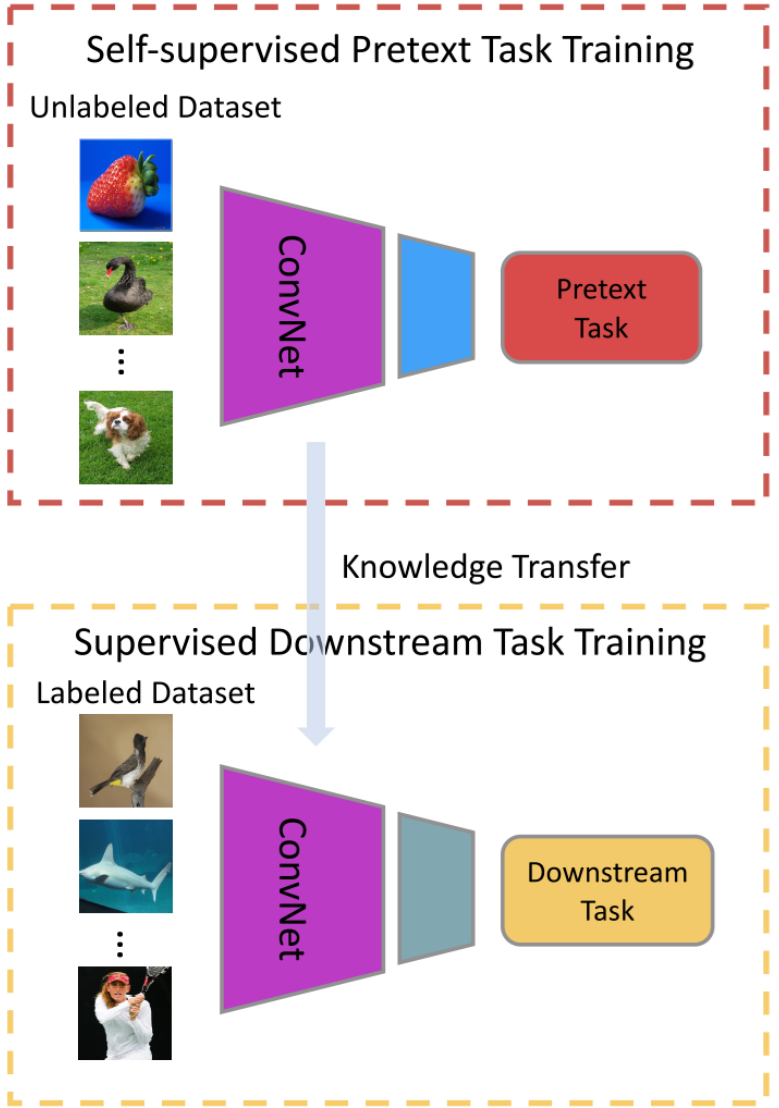
Deep Learning

Mohammad Reza Mohammadi

2021

Self-supervised learning

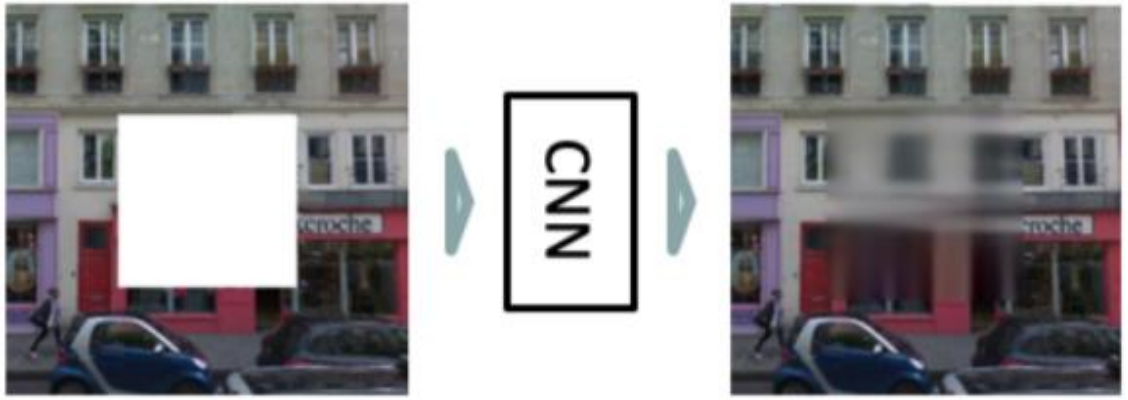
- Self-supervised learning methods are proposed to learn general features from large-scale unlabeled data without using any human-annotated labels
- The pretext tasks share two common properties:
  - Visual features need to be captured by ConvNets to solve the pretext tasks
  - Pseudo labels for the pretext task can be automatically generated based on the attributes of images or videos

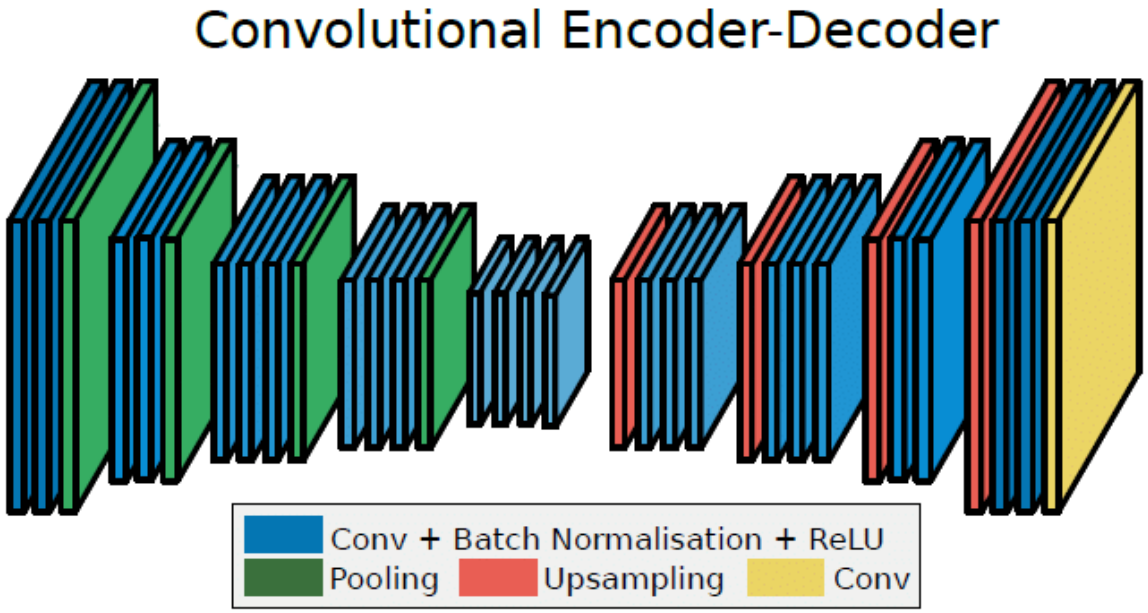


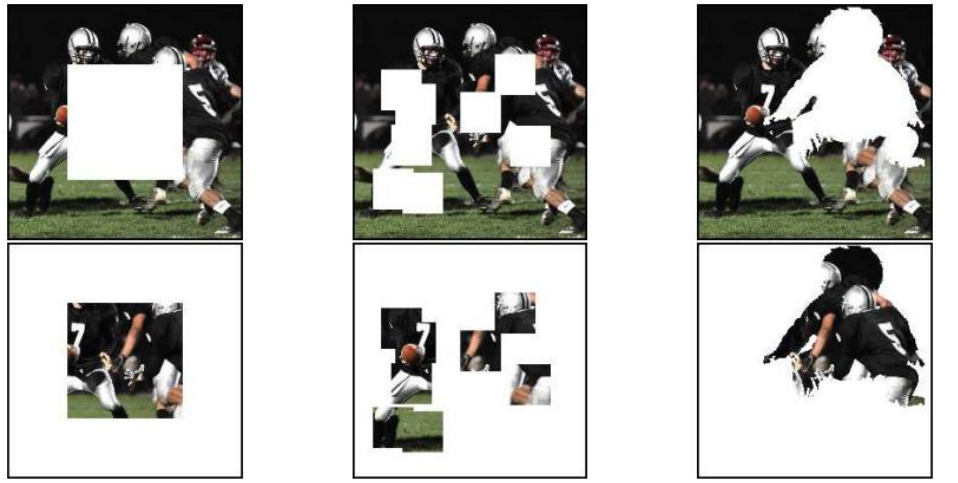
L. Jing, and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey." *IEEE Transactions on PAMI* (2020).

Example pretext task: context prediction

$$\mathcal{L}_{rec}(x) = \left\| \widehat{M} \odot \left(x - F\left(\left(1 - \widehat{M} \right) \odot x \right) \right) \right\|_2$$





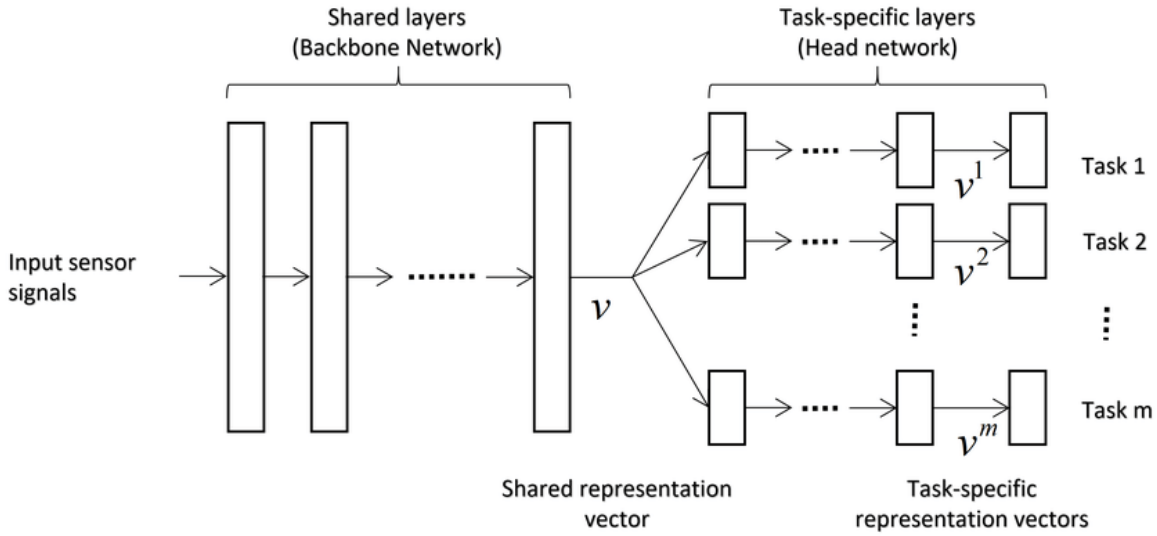


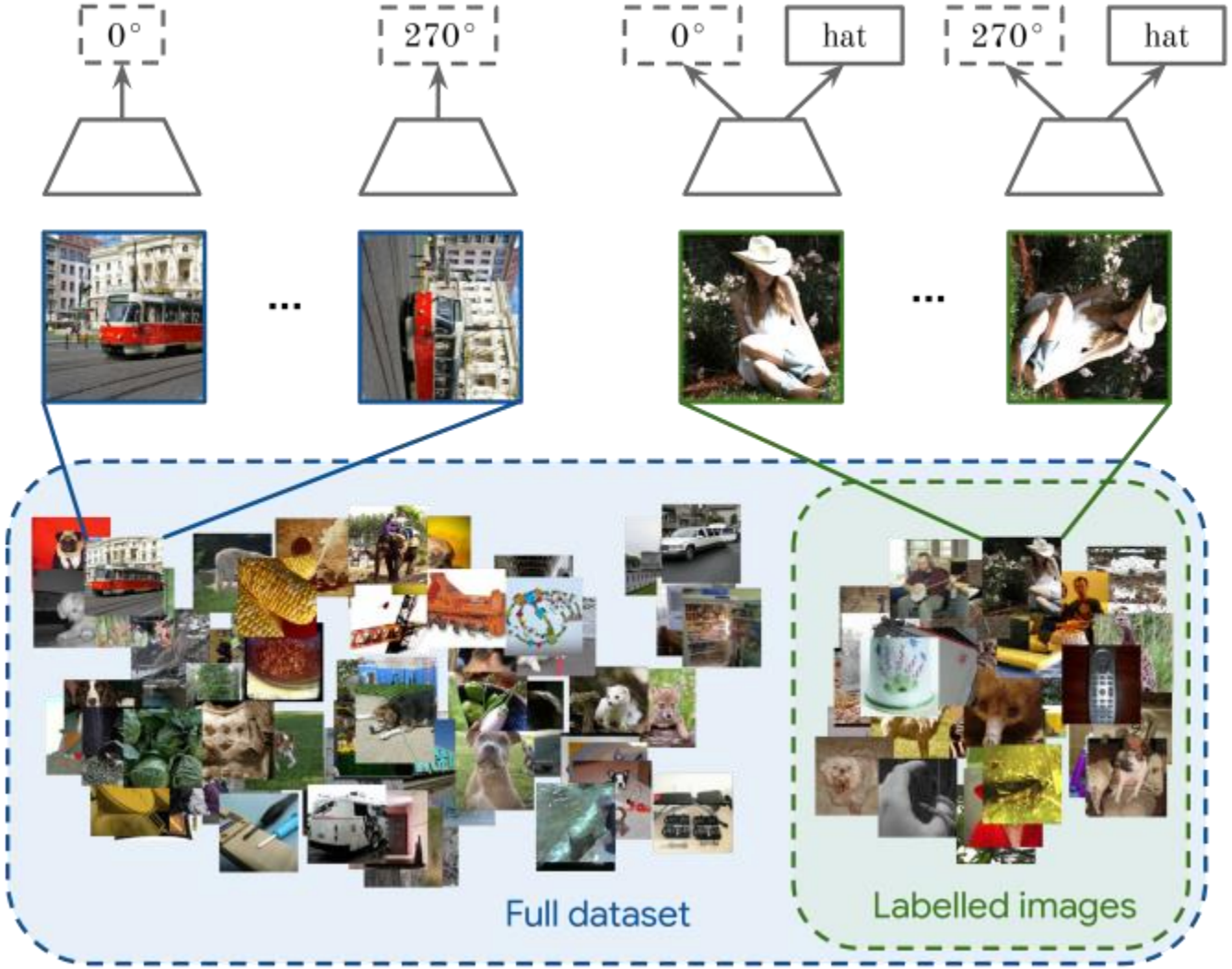
(a) Central region (b) Random block (c) Random region

D. Pathak, et al. "Context encoders: Feature learning by inpainting." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Multi-task learning

- The results show that combining tasks (even via a naive multihead architecture) always improves performance

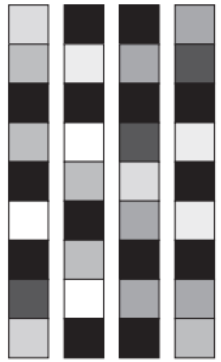




X. Zhai, et al. "S4l: Self-supervised semi-supervised learning." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

Word embeddings

- Word embeddings pack more information into far fewer dimensions
- They can be pre-trained on large amounts of text training data



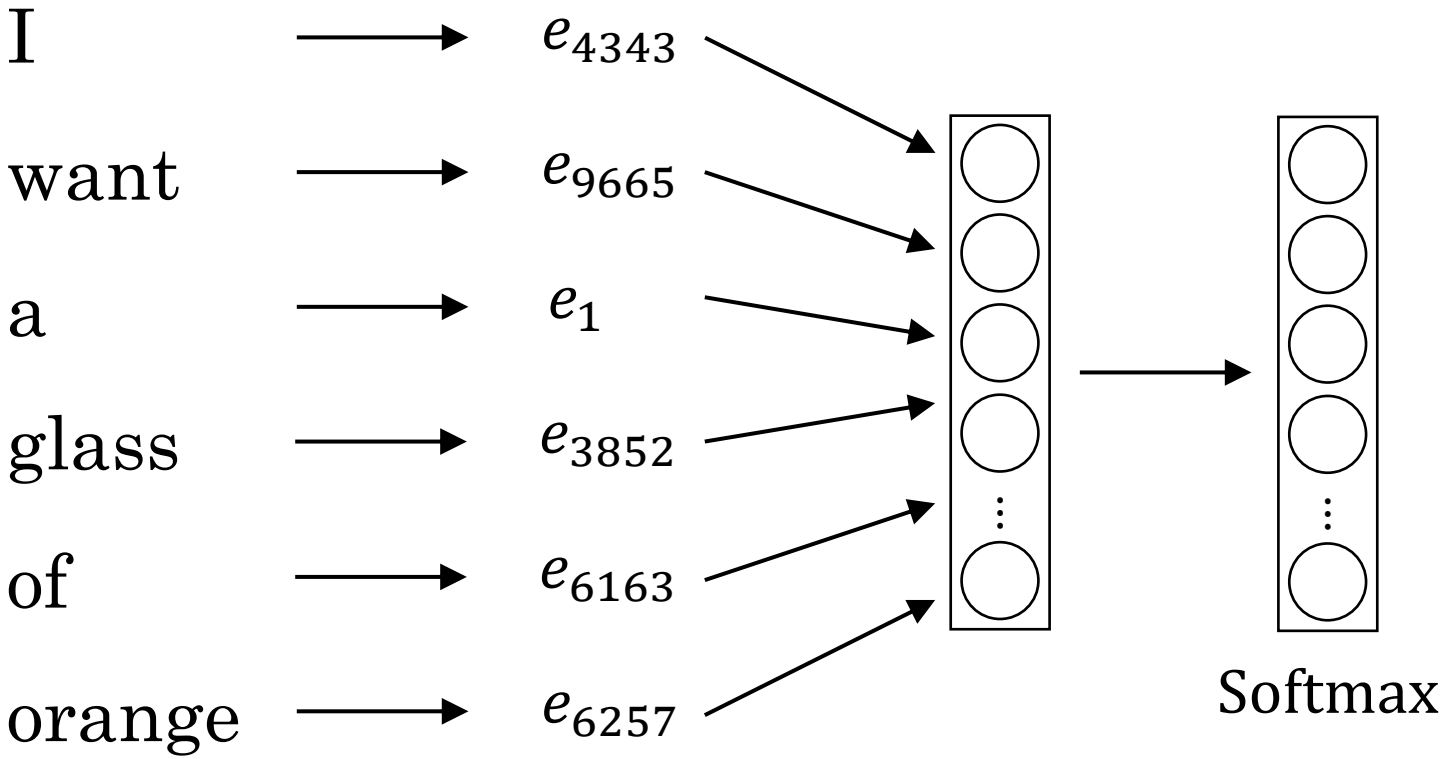
Word embeddings:
- Dense
- Lower-dimensional
- Learned from data

One-hot word vectors:
- Sparse
- High-dimensional
- Hardcoded

Neural language model

I want a glass of orange _____.

4343 9665 1 3852 6163 6257



Bengio, Yoshua, et al. "A neural probabilistic language model." *The journal of machine learning research*, 2003.

Other context/target pairs

- Example sentence:
  - I want a glass of orange juice to go along with my cereal.

- Context
  - Last 4 words
  - 4 words on left & right
  - Last 1 word
  - Nearby 1 word

Word2Vec

Context	Target

- Skip-grams
  - rather than having the context be always the last N words
  - randomly pick a word to be the context word
  - randomly pick another word within some window
  - set up a supervised learning problem where:
    - given the context word
    - predict what is a randomly chosen word

I want a glass of orange juice to go along with my cereal.

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

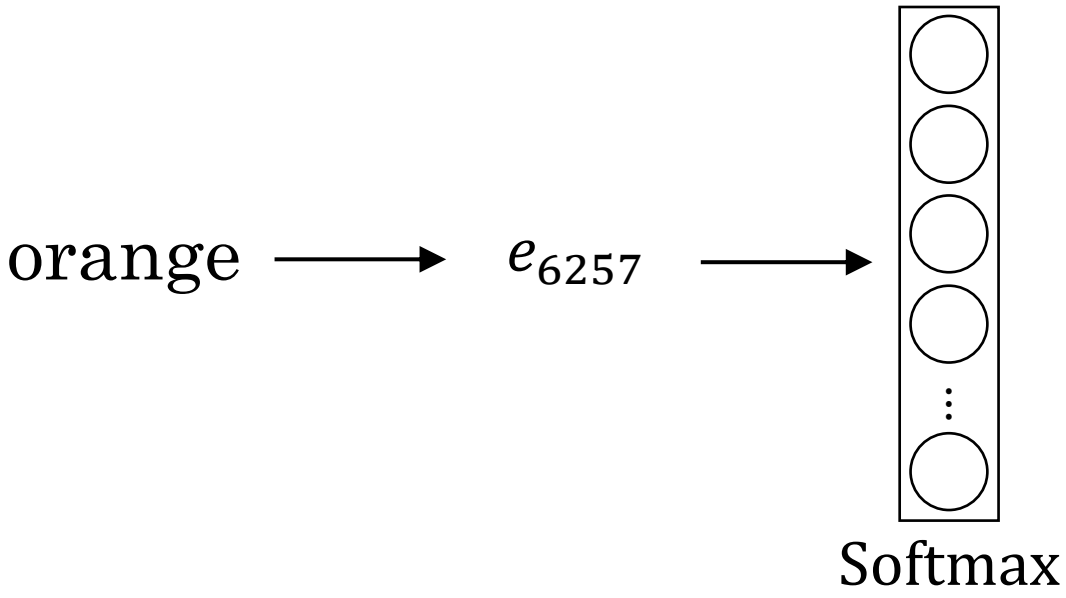
Word2Vec

- Vocab size: 10,000

Context c ("orange") ⇒ Target t ("juice")

6257 4834

Context	Target
orange	juice
orange	glass
orange	with



$$\mathcal{L}_{ce} = -\sum_{j=1}^{10,000} y_j \log \hat{y}_j$$

$$p(t|c) = \frac{e^{w_t^T e_c + b_t}}{\sum_{j=1}^{10,000} e^{w_j^T e_c + b_j}}$$

Word2Vec

- Problems with Softmax classification
  - The primary problem is the computational speed
  - Especially for very large vocabularies (e.g., size 100,000 or 1,000,000)

$$p(t|c) = \frac{e^{w_t^T e_c + b_t}}{\sum_{j=1}^{10,000} e^{w_j^T e_c + b_j}}$$

Word2Vec

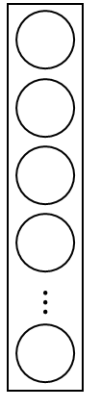
- Negative sampling
  - Create a new supervised learning problem
  - Given a pair of words
    - predict that is a context-target pair?
  - So the problem is really given a pair of words, do you think they appear together?
  - recommend that maybe k is:
    - 5 to 20 for smaller data sets
    - 2 to 5 for larger data sets

Context	Word	Target

I want a glass of orange juice to go along with my cereal.

Word2Vec

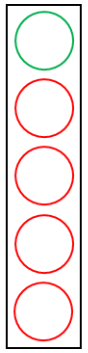
- Negative sampling

orange → e_{6257} → 

Softmax

$$p(t|c) = \frac{e^{w_t^T e_c + b_t}}{\sum_{j=1}^{10,000} e^{w_j^T e_c + b_j}}$$

Context	Word	Target
orange	juice	1
orange	king	0
orange	book	0
orange	the	0
orange	of	0

orange → e_{6257} → 

Sigmoid

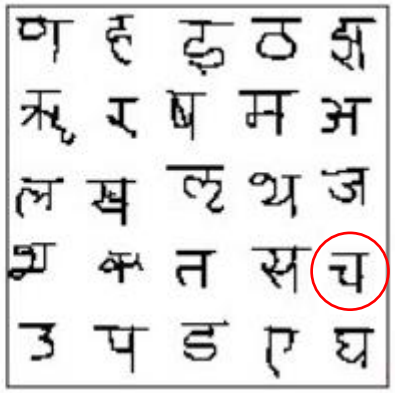
$$p(y = 1|t, c) = \sigma(w_t^T e_c + b_t)$$

One shot learning

- One-shot learning is a classification task in which one example for each class is used to classify many new examples

Test Image

Support Set



https://towardsdatascience.com/one-shot-learning-with-siamese-networks-using-keras-17f34e75bb3d

One shot face recognition

- Learning from example to recognize the person again
  - Is a ConvNet + Softmax well for one shot face recognition?
    - Not enough data to train a robust neural network
    - What if a new person joins our team?
  - Instead, we are going to learn a “similarity” function