Deep Learning 4

Autoencoder, Attention (spatial transformer), Multi-modal learning, Neural Turing Machine, Memory Networks, Generative Adversarial Net

Autoencoder

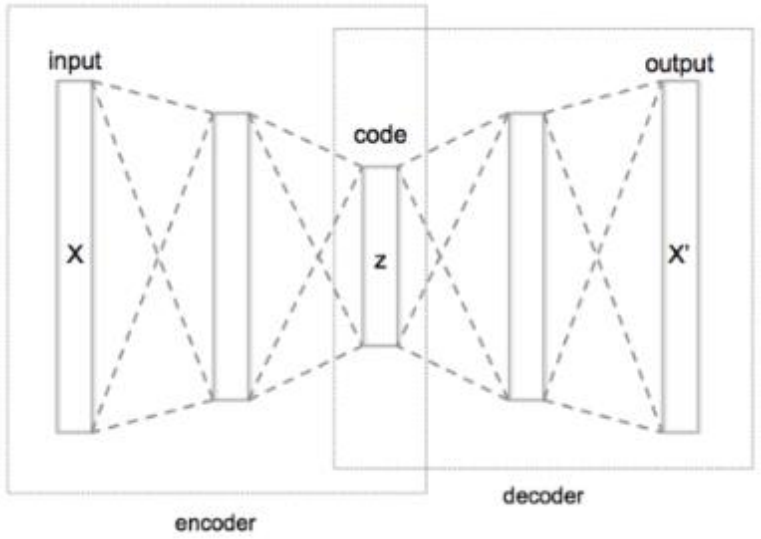
Autoencoder

-
  -
    -
    -
  -

$$\phi : \mathcal{X} \to \mathcal{F}$$

$$\psi : \mathcal{F} \to \mathcal{X}$$

$$\arg\min_{\phi,\psi} \|X - (\psi \circ \phi)X\|^2$$



Denoising Autoencoder

-
  -
  -

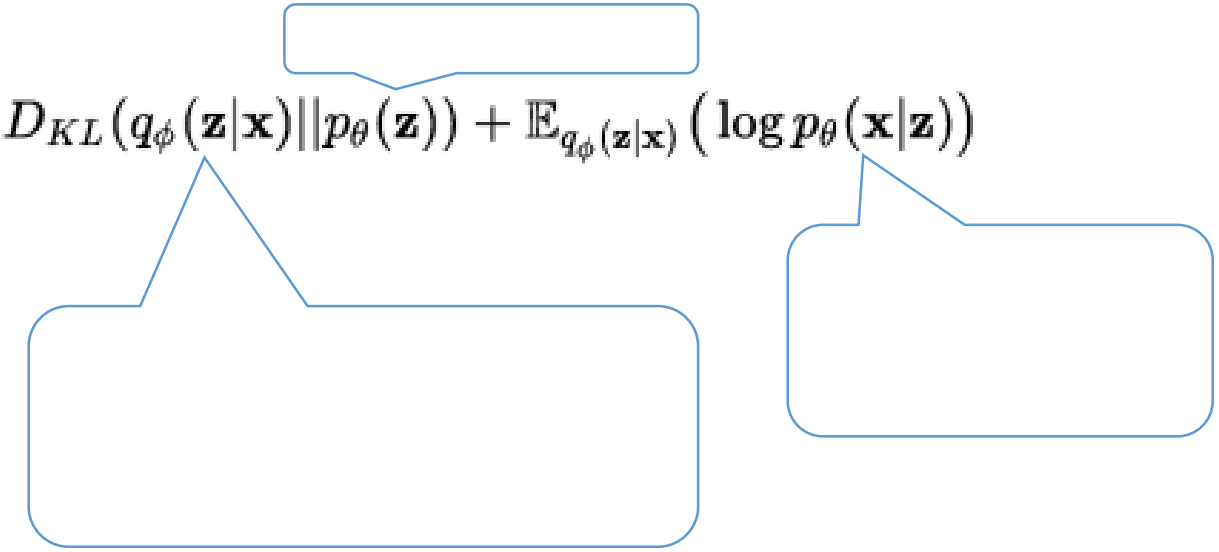
Sparse Autoencoder

- **sparsity**
  - $$J_{\text{sparse}}(W,b) = J(W,b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho||\hat{\rho}_j),$$
  - ” Ḋ
  -

Variational autoencoder (VAE)

-
-

$$\mathcal{L}(\phi, \theta, \mathbf{x}) = -D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\big(\log p_\theta(\mathbf{x}|\mathbf{z})\big)$$



A quick intro to variational inference

•
•

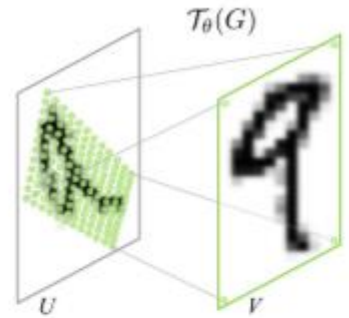
$$q^*(\mathbf{z}) = \arg\min_{q(\mathbf{z})\in\mathcal{Q}} \mathrm{KL}\left(q(\mathbf{z})\|p(\mathbf{z}\,|\,\mathbf{x})\right).$$

~ Æ • } • v [š ‰ v } v Ì •

A small
Localization network
predicts transform θ

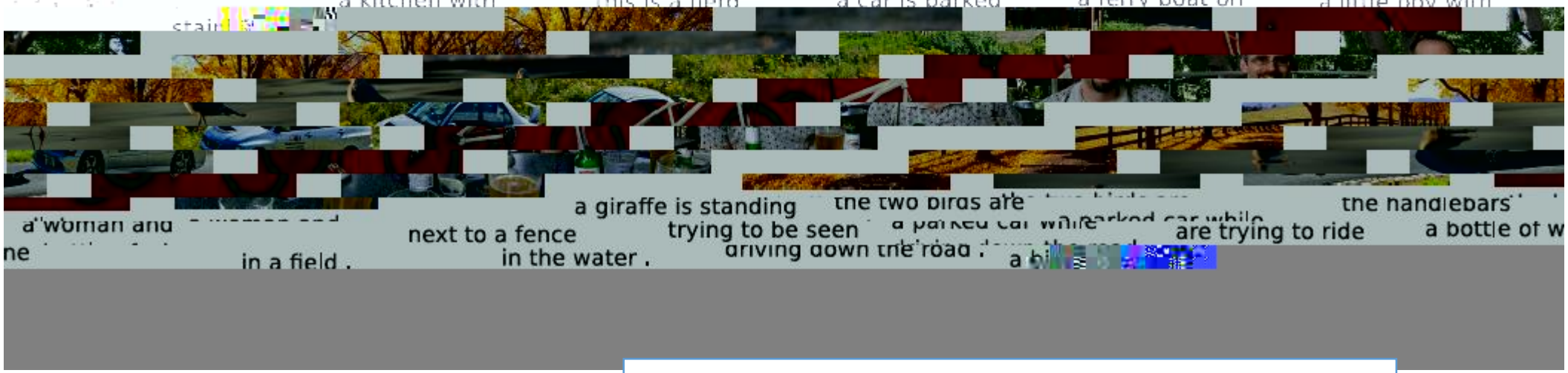
Grid generator uses θ to compute sampling grid

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

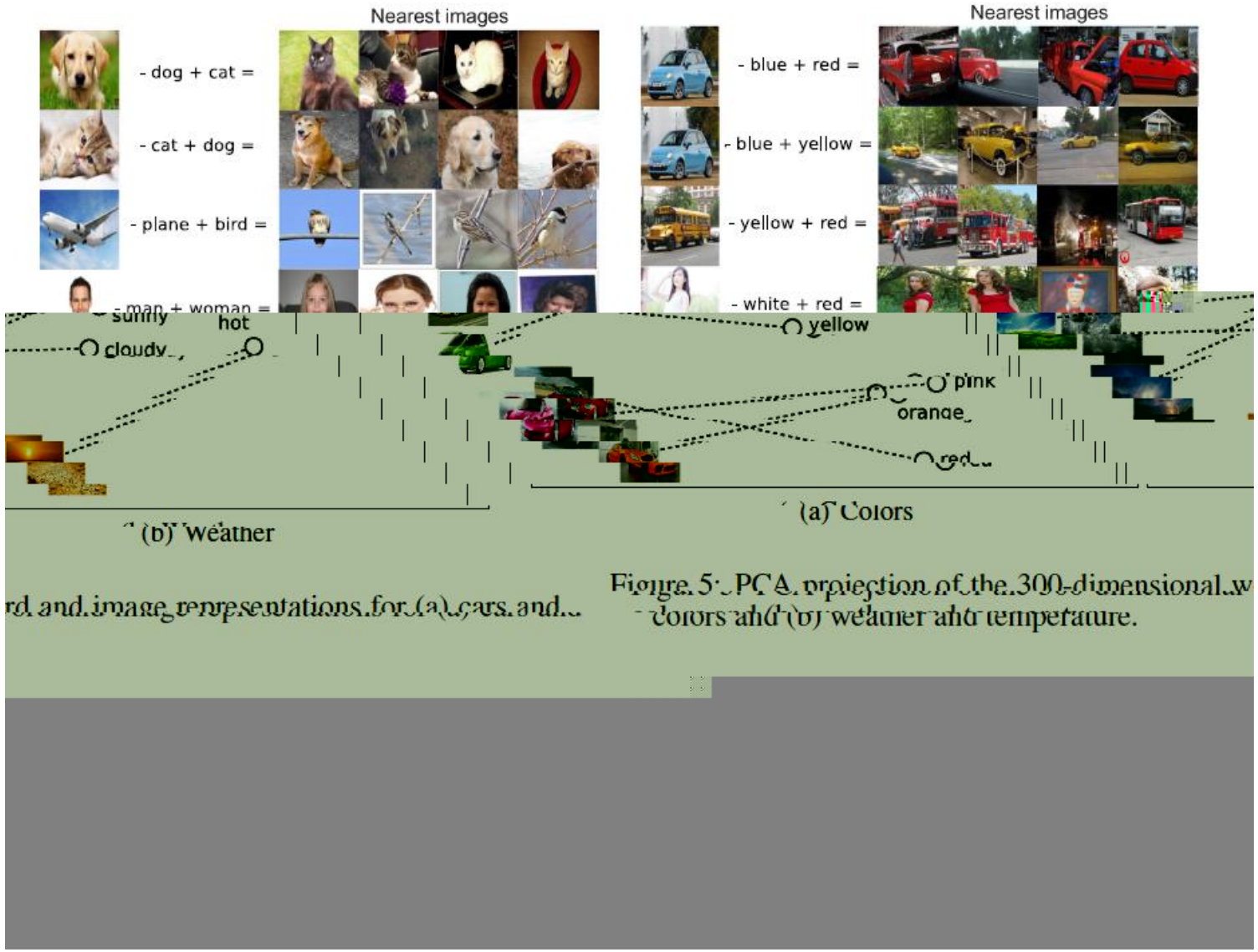


—

(ώ̣̈ȟώ̣̈) ᴺ ˝Ɣ̣ ˜O



(ώ̣̈ȟώ̣̈) ᴺ ˝Ɣ̣ ˜O



rd and image representations for (a) cars and...

Figure 5: PCA projection of the 300-dimensional w
- colors and (b) weather and temperature.

Details

Let $\mathbf{q} \in \mathbb{R}^D$ denote an image feature vector

- $\mathbf{x} = \mathbf{W}_I \cdot \mathbf{q} \in \mathbb{R}^K$ be the image embedding
- image description $S = \{w_1, \ldots, w_N\}$ with words $w_1, \ldots, w_N$

[redacted]

ώ ı
Á } Œ î À ~ Á } v [š o Œ v

Details

- —ɖ ώ ĩ

$$\min_{\boldsymbol{\theta}} \sum_{\mathbf{x}} \sum_{k} \max\{0, \alpha - s(\mathbf{x}, \mathbf{v}) + s(\mathbf{x}, \mathbf{v}_k)\} + \sum_{\mathbf{v}} \sum_{k} \max\{0, \alpha - s(\mathbf{v}, \mathbf{x}) + s(\mathbf{v}, \mathbf{x}_k)\}$$

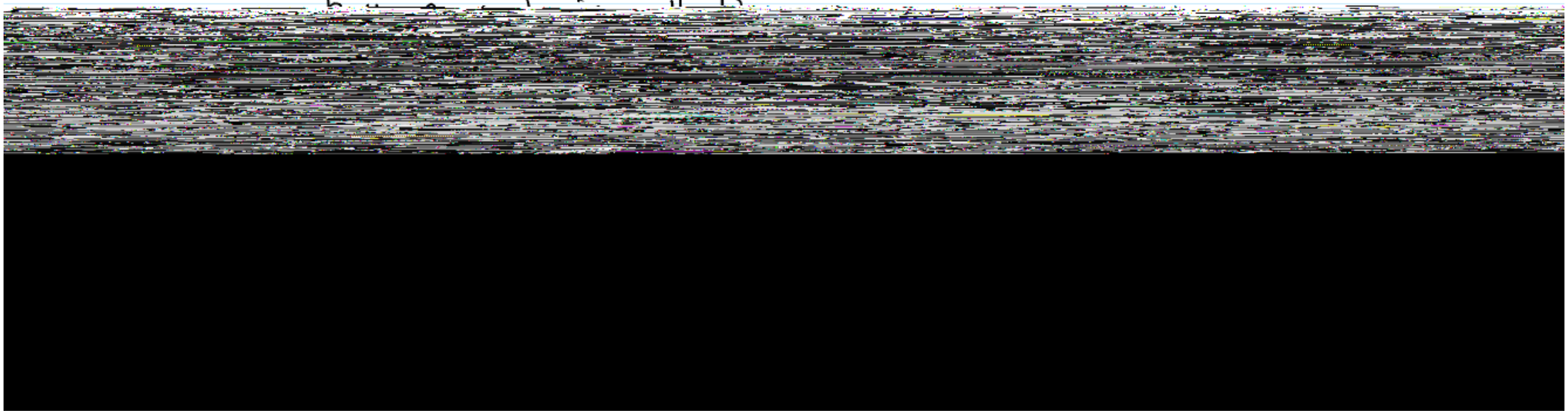
scoring function $s(\mathbf{x}, \mathbf{v}) = \mathbf{x} \cdot \mathbf{v}$.
$\mathbf{v}_k$ is a contrastive (non-descriptive) sentence for image embedding $\mathbf{x}$, and vice-versa with $\mathbf{x}_k$.

Neural Turing Machine [Graves et al.]

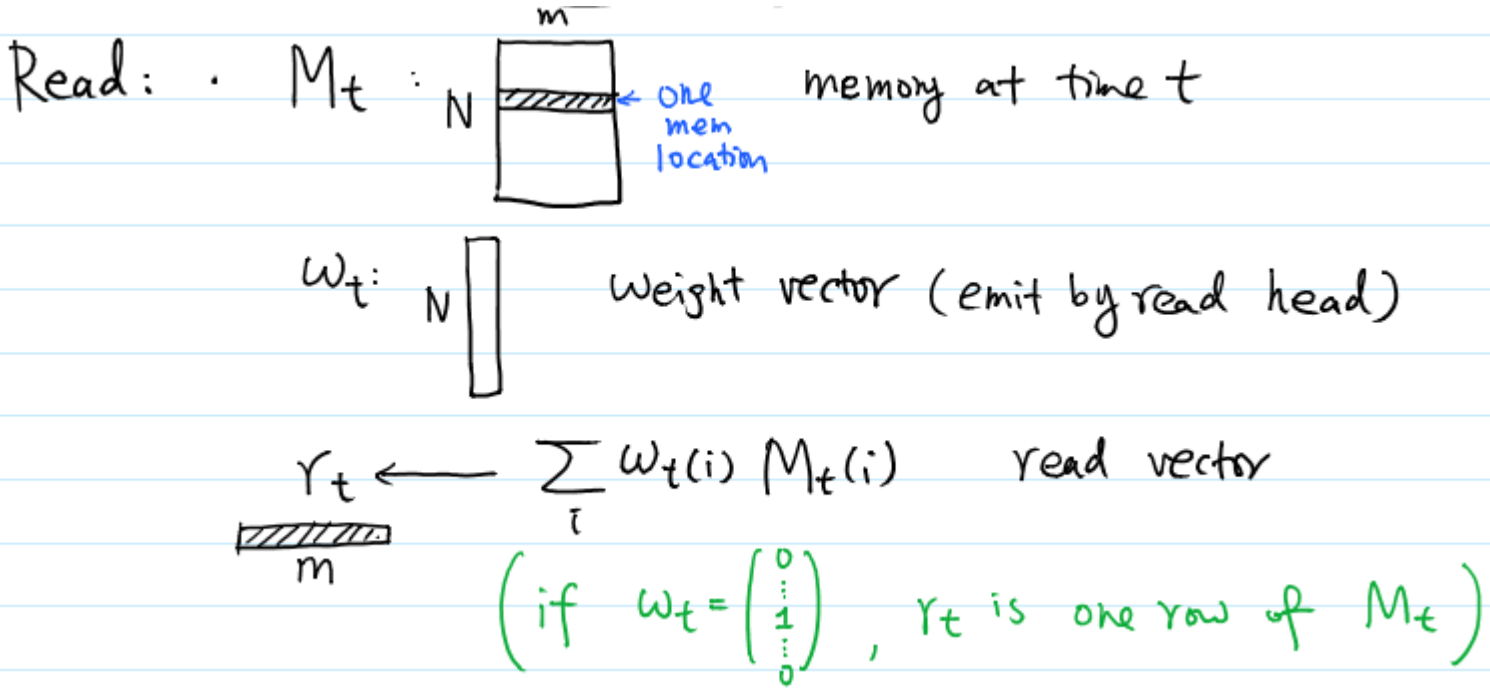
“Memory”

concept of working memory.

— in psychology: explain short term manipulation of information



Read


Read:
M_t :
m
N
one mem location
memory at time t
w_t :
N
weight vector (emit by read head)
$r_t \leftarrow \sum_i w_t(i)\, M_t(i)$
read vector
m
(if $w_t = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, r_t is one row of M_t)

Write

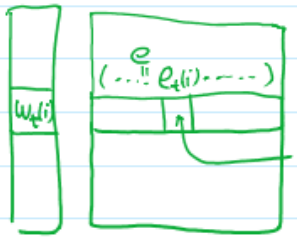
Write: first erase, then add.

w_t: N-dimensional vector — weight vector (emit by write head)

Erase:

$$\tilde{M}_t(i) \leftarrow M_{t-1}(i) \cdot [\mathbf{1} - w_t(i)\, e_t]$$

- $\cdot$: pointwise
- $\mathbf{1}$: $(1, 1, \ldots, 1)$, m
- e_t: erase vector, $(0.1, 0, 0, 0.99, \ldots)$, m, in $(0, 1)$



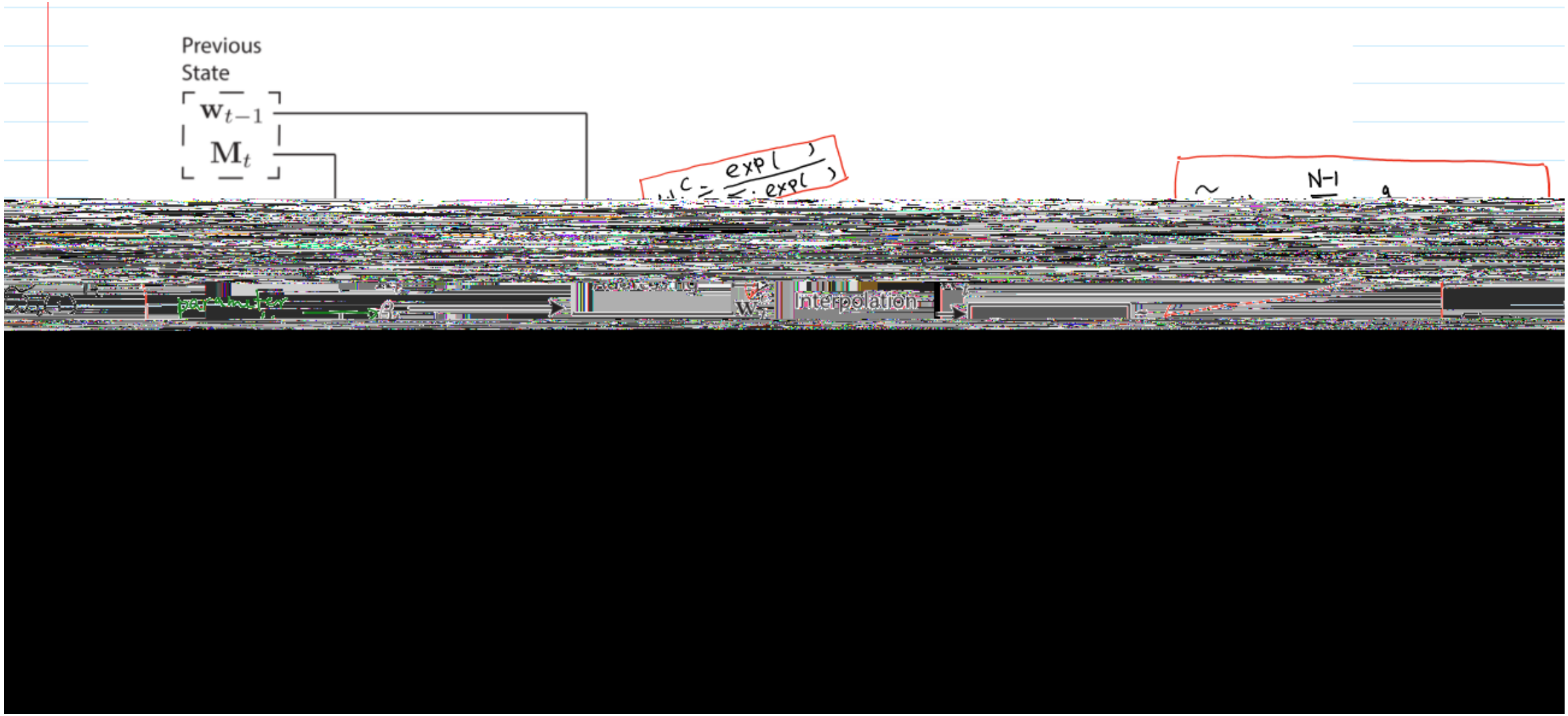
it is erased if $w_t(i)$ and $e_t(i)$ are 1

Add:

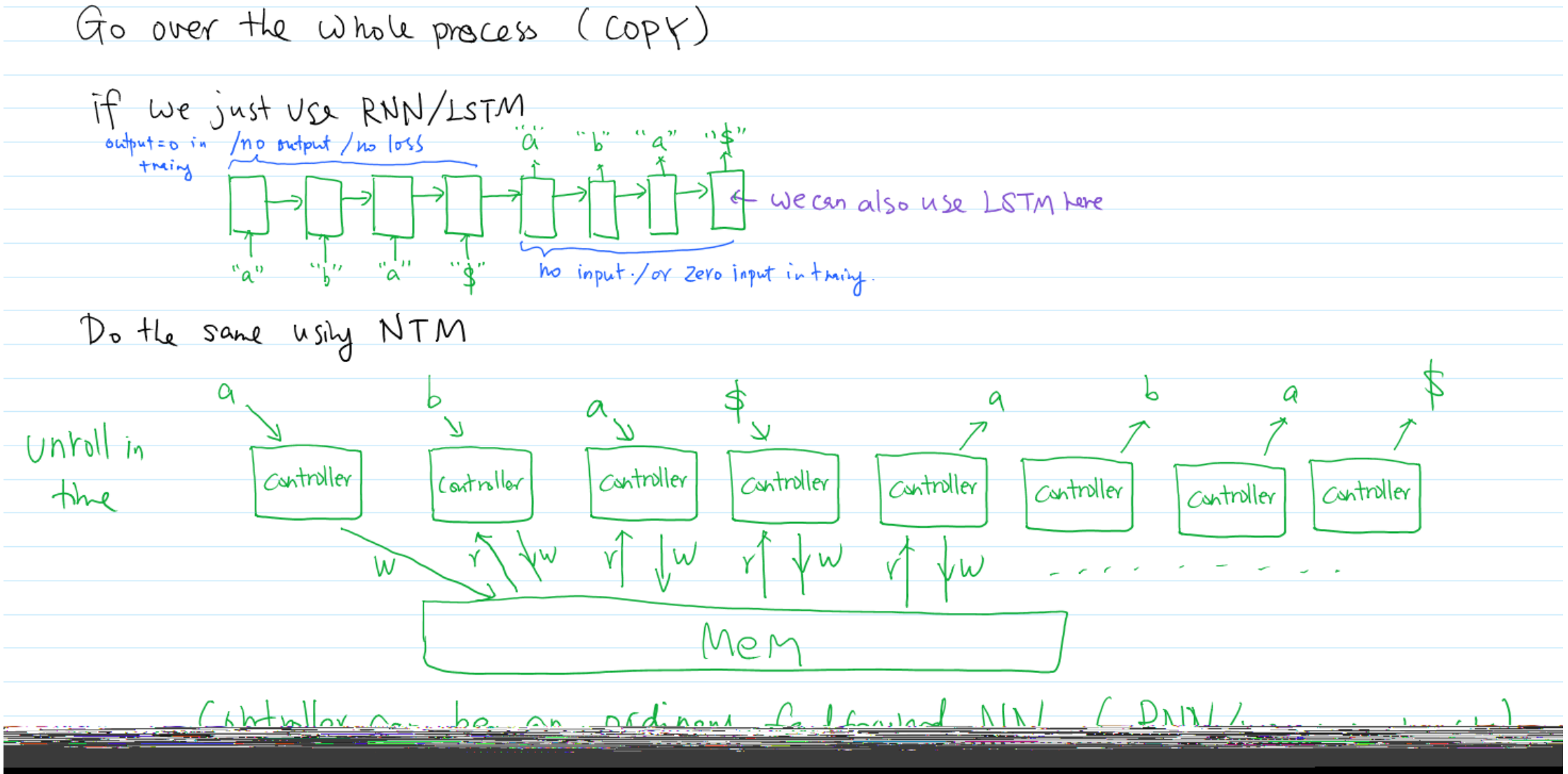
$$M_t(i) \leftarrow \tilde{M}_t(i) + w_t(i)\, a_t$$

- a_t: add vector, $(\quad)$, m

Addressing Mechanism (overview)



Go over the process



t the bottom

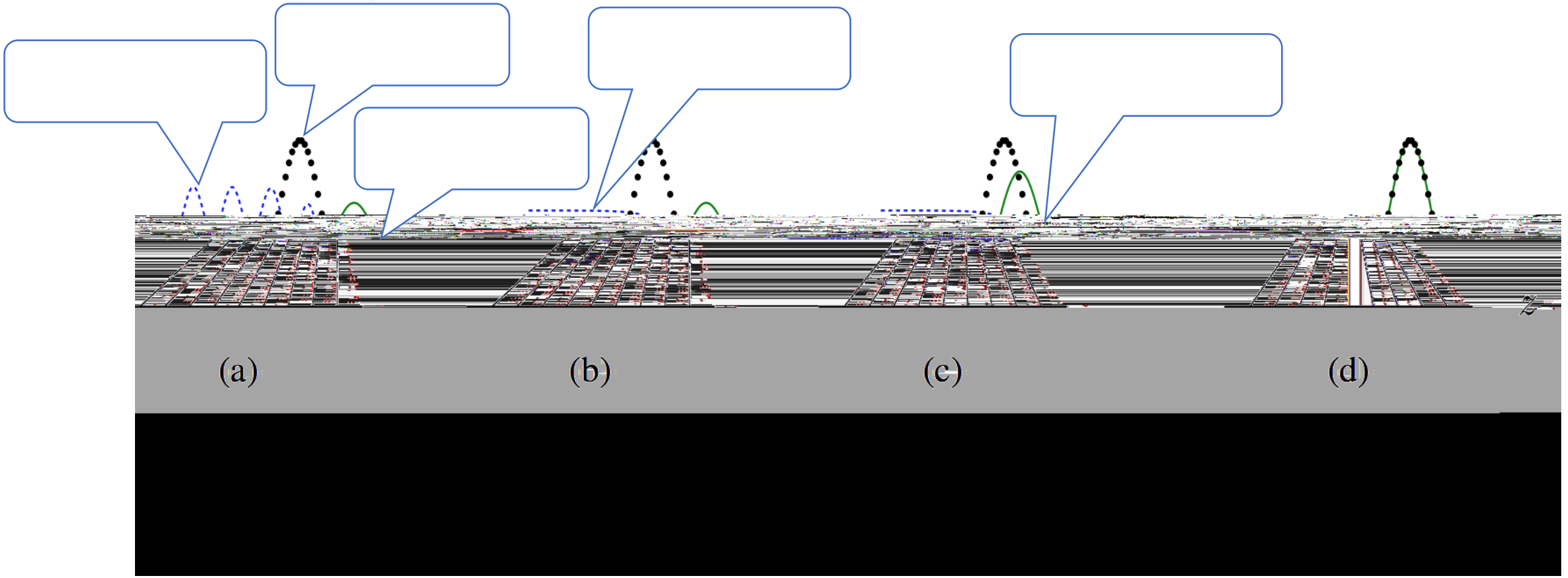
the red arrow

Memory Network

[Weston et al.][Sukhbaatar et al.]

Memory Network [Weston et al.][Sukhbaatar et al.]

Training Process



Another visualized training process, see https://cs.stanford.edu/people/karpathy/gan/

Experiments

Model	MNIST	TFD
DBN [3]	138 ± 2	1909 ± 66
Stacked CAE [3]	121 ± 1.6	$\mathbf{2110 \pm 50}$
Deep GSN [5]	214 ± 1.1	1890 ± 29
Adversarial nets	$\mathbf{225 \pm 2}$	$\mathbf{2057 \pm 26}$

Final Notes

- t Z À }À Œ šZ •] •U v • À Œ o Œ vš ^
-
-
  - µš]š } •v[š u v šZ š >]• iµ•š ^šµv]vP ‰ Œ u š Œ•
- /u‰}Œš vš dZ]vP• t] v[š }À Œ
    -
    -
    -
    - ^š l]vP šŒ]š]}v o ^•Z oo}Á_ u} o• YY
    -
    -
    -
    -
    - YXX

-
  -
  -
  - **Deep Atari Deep Learn**
  - t
    -
    -

- ^š v(}Œ U •o] • (}Œ ^
  D u}ŒÇ E šÁ}Œl•_ Ç
-
-