



Attention in Deep Models

(Sept 28, 2016)

YANG Jiancheng



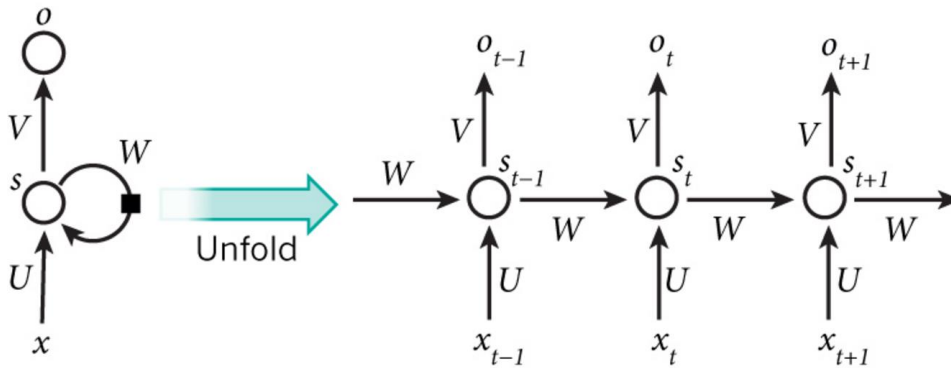
Outline

- **I. Attention in Sequence Models**
- **II. Attention in CN based Models**
- **III. Spatial Transformer Network**

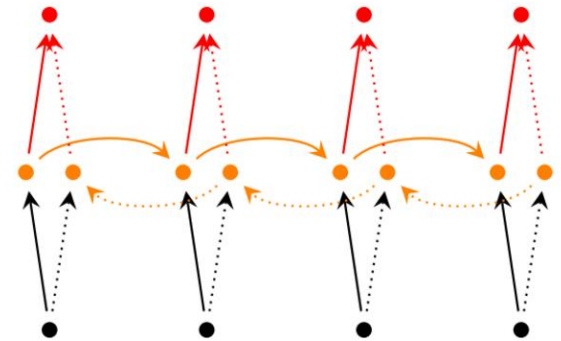


• I. Attention in Sequence Models

Normal RNN



RNN and the unfolding

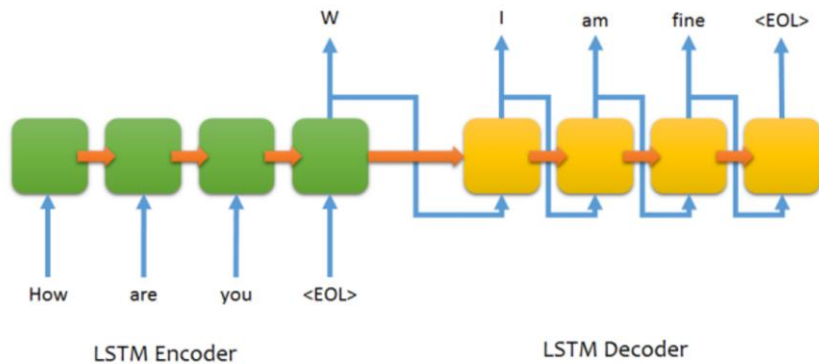


Bidirectional RNN

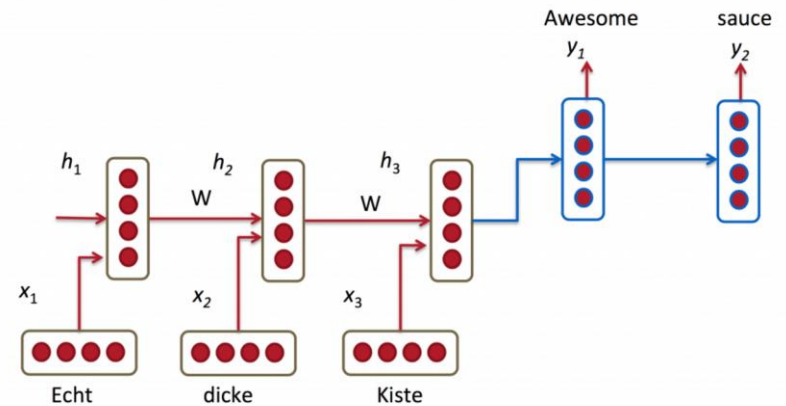


• I. Attention in Sequence Models

Seq2Seq and Machine Translation



Seq2seq



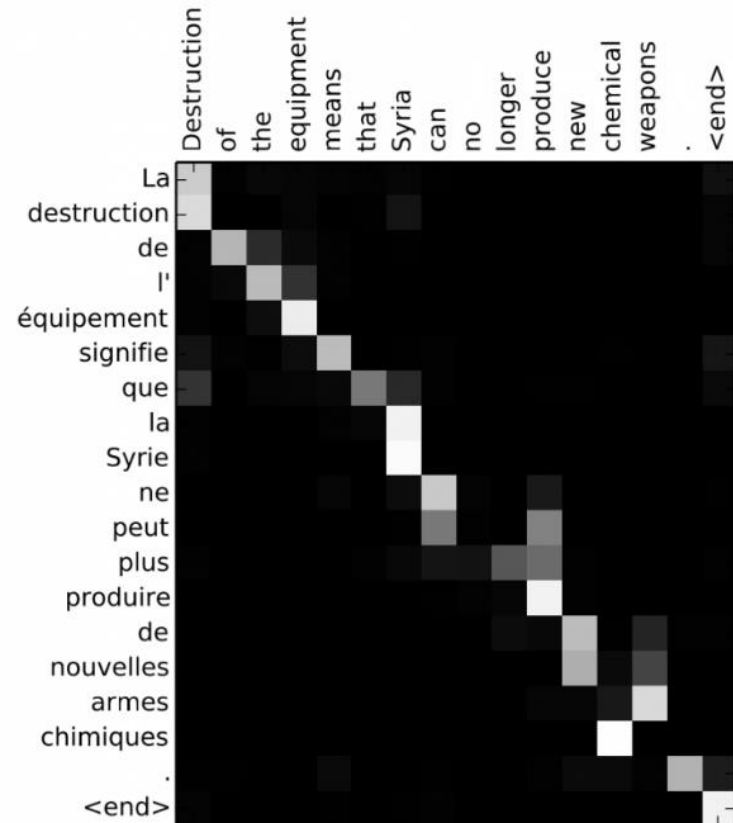
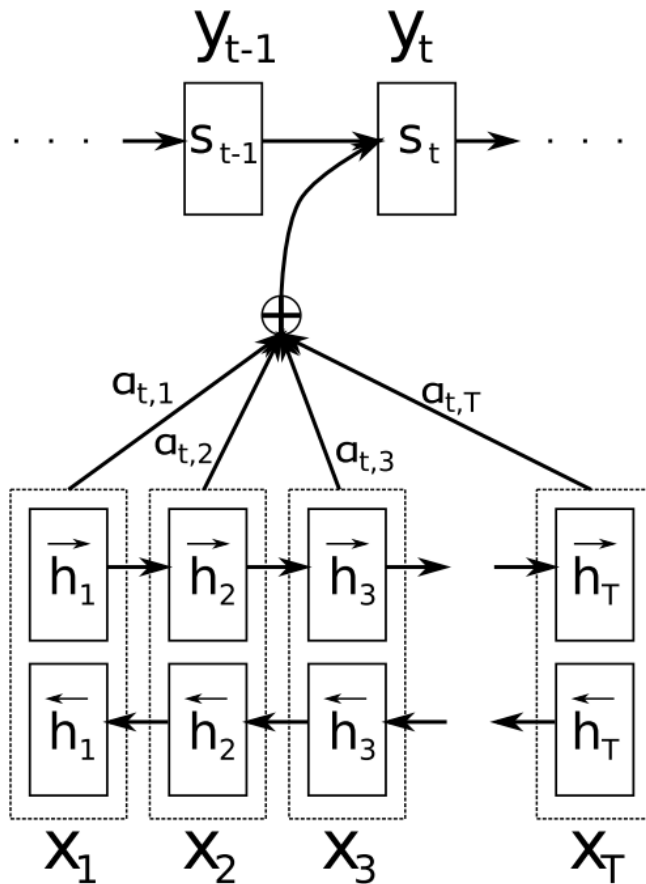
Neural Machine Translation

Sequence to Sequence Learning with Neural Networks



• I. Attention in Sequence Models

Attention Based RNN in Neural Machine Translation



Neural Machine Translation by Jointly Learning to Align and Translate



• I. Attention in Sequence Models

Attention Based Article Understanding

by *ent423* ,*ent261* correspondent updated 9:49 pm et ,thu march 19 ,2015 (*ent261*) a *ent114* was killed in a parachute accident in *ent45* ,*ent85* ,near *ent312* ,a *ent119* official told *ent261* on wednesday .he was identified thursday as special warfare operator 3rd class *ent23* ,29 ,of *ent187* ,*ent265* .`` *ent23* distinguished himself consistently throughout his career .he was the epitome of the quiet professional in all facets of his life ,and he leaves an inspiring legacy of natural tenacity and focused

...

ent119 identifies deceased sailor as **X** ,who leaves behind a wife

by *ent270* ,*ent223* updated 9:35 am et ,mon march 2 ,2015 (*ent223*) *ent63* went familial for fall at its fashion show in *ent231* on sunday ,dedicating its collection to ``mamma " with nary a pair of ``mom jeans " in sight .*ent164* and *ent21* ,who are behind the *ent196* brand ,sent models down the runway in decidedly feminine dresses and skirts adorned with roses ,lace and even embroidered doodles by the designers ' own nieces and nephews .many of the looks featured saccharine needlework phrases like ``i love you ,

...

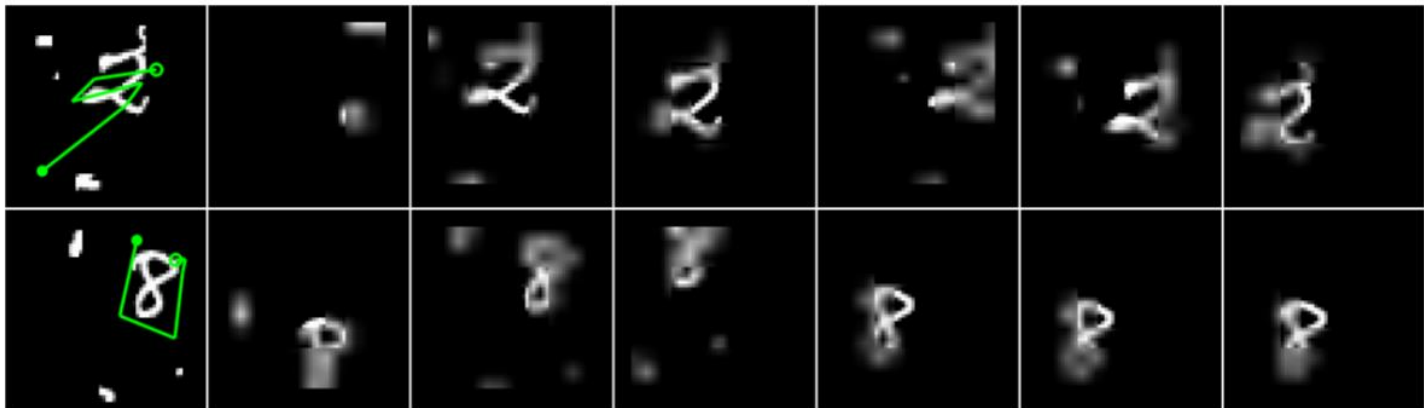
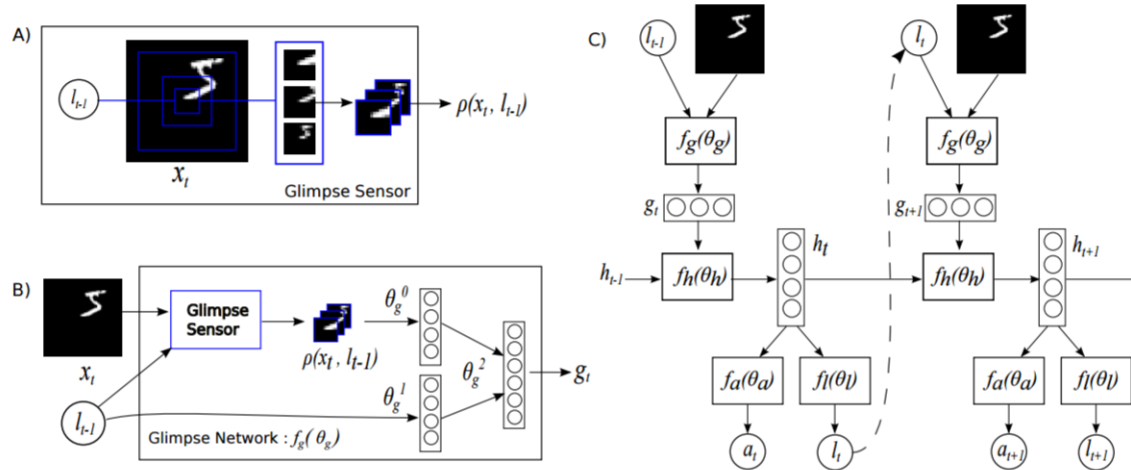
X dedicated their fall fashion show to moms

Teaching Machines to Read and Comprehend



• II. Attention in CN based Models

Attention with Reinforcement Learning

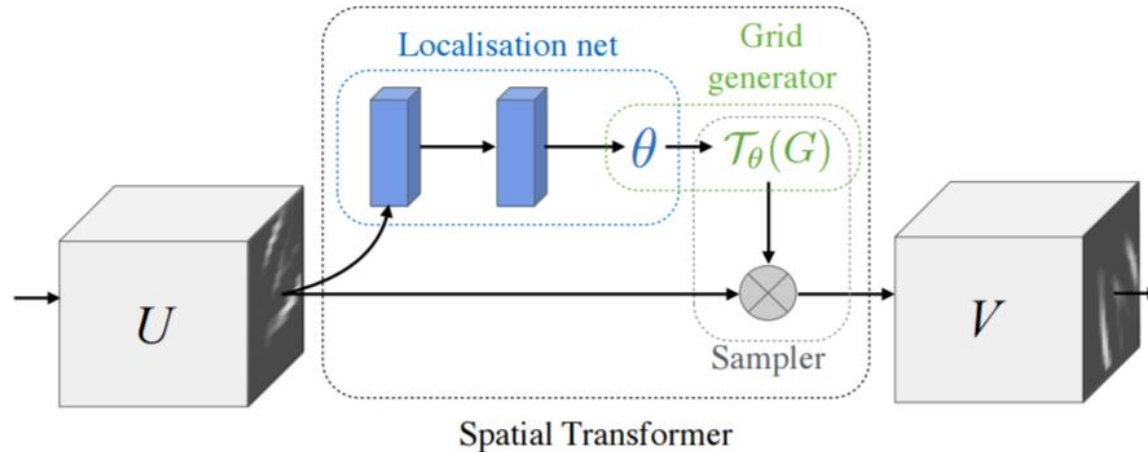


Recurrent Models of Visual Attention



• II. Attention in CN based Models

Attention with differentiable module - STN



Spatial Transformer Networks



• III. Spatial Transformer Network

State-of-the-art performance (GTSRB dataset)



IDSIA 99.46% top-1 accuracy.

Moodstocks 99.61% top-1 accuracy with a much simpler pipeline.

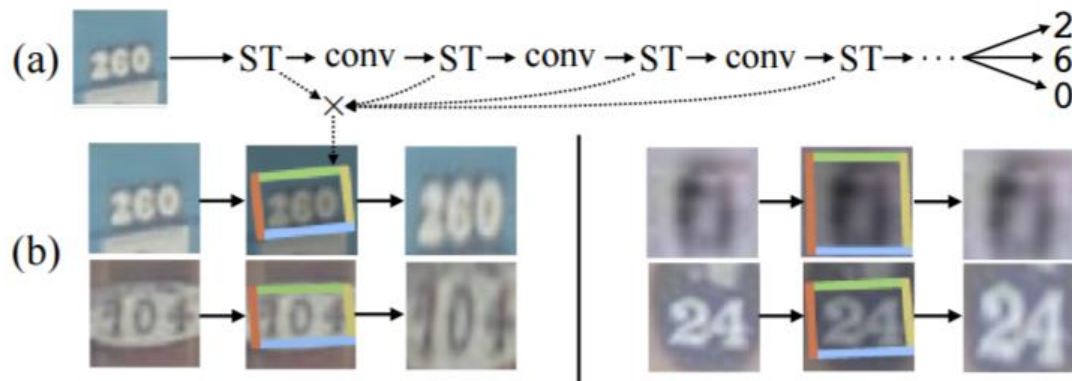
Pipeline	IDSIA	Moodstocks
Augmentations	Yes (i)	No
Jittering	Yes (ii)	No
Network	~90M weights (iii)	~20M weights (iv)



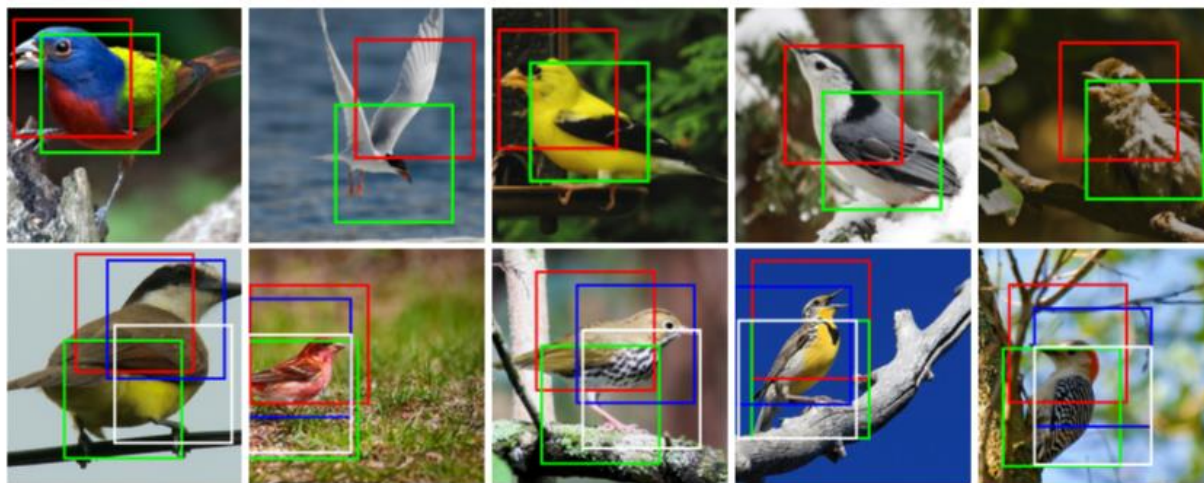
• III. Spatial Transformer Network

State-of-the-art performance (SVHN & CUB-200-2011)

Model	Size	
	64px	128px
Maxout CNN [13]	4.0	-
CNN (ours)	4.0	5.6
DRAM* [1]	3.9	4.5
ST-CNN	3.7	3.9
	3.6	3.9



Model		
Cimpoi '15 [5]		66.7
Zhang '14 [40]		74.9
Branson '14 [3]		75.7
Lin '15 [23]		80.9
Simon '15 [30]		81.0
CNN (ours) 224px		82.3
2×ST-CNN 224px		83.1
2×ST-CNN 448px		83.9
4×ST-CNN 448px		84.1





• III. Spatial Transformer Network

My trial (12 animals classification)



More on PawNation: Attack, Awesome, Reptiles, Scary, Science, Wild Animals



By Paul Ciampelli (Sci)

6. Hyena (1,100 psi)

Hyenas suffer from a cowardly scavenger "The Lion King." right Although hyenas are hey, so are dogs — food the majority of t and a biting force the powerful of all mam



Single ResNet-50: ~85%
STN + ResNet-50: ~90%



• III. Spatial Transformer Network

Unreasonable effectiveness

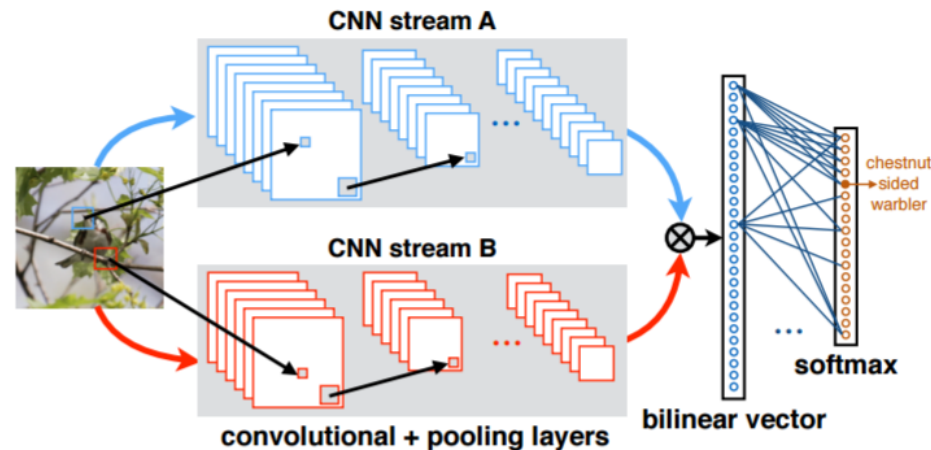
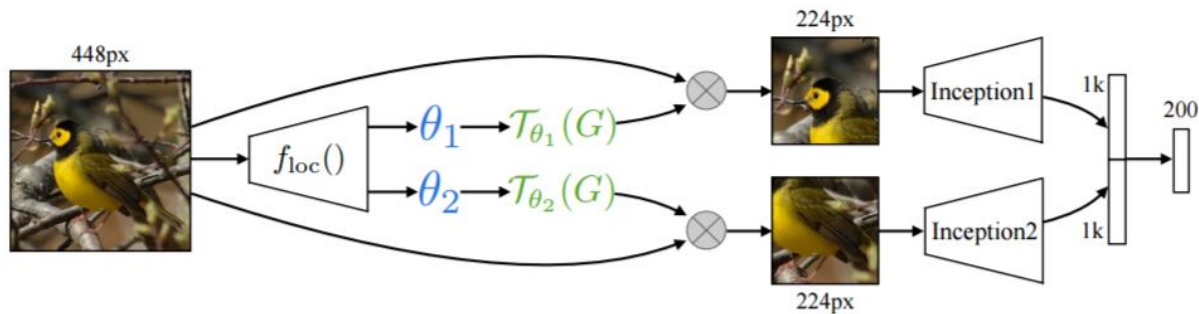




• III. Spatial Transformer Network

Fine-grained Visual Recognition

CUB-200-2011

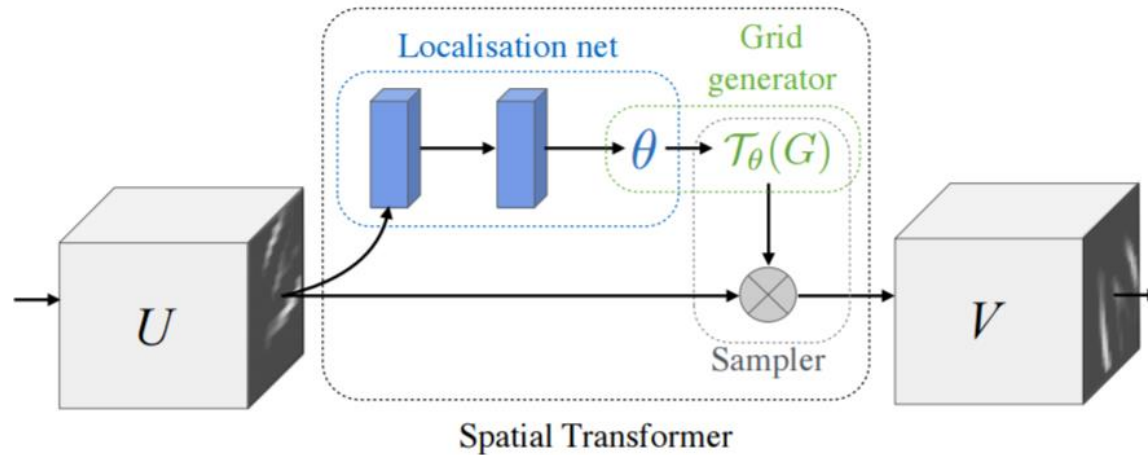


Bilinear CNN Models for Fine-grained Visual Recognition



• III. Spatial Transformer Network

Layer Structure

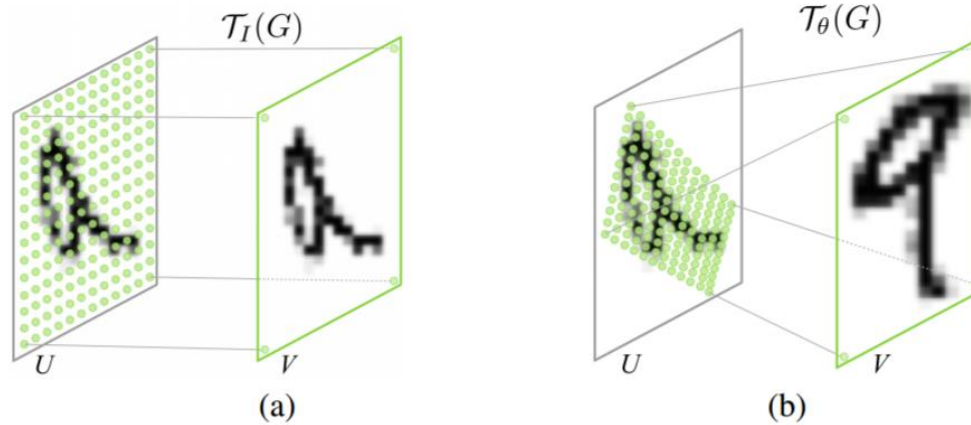


1. Localisation Network
2. Parameterised Sampling Grid
3. Differentiable Image Sampling



• III. Spatial Transformer Network

Parameterised Sampling Grid



$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \mathcal{T}_\theta(G_i) = \mathbf{A}_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$



Bibliography

- [1] [Attention and Memory in Deep Learning and NLP](#)
- [2] [Sequence to Sequence Learning with Neural Networks](#)
- [3] [Neural Machine Translation by Jointly Learning to Align and Translate](#)
- [4] [Recurrent Models of Visual Attention](#)
- [5] [Spatial Transformer Networks](#)



Thanks for your **attention!**

