Lecture: Deep Learning and Differential Programming

3.2 Visualization of learned knowledge

https://liris.cnrs.fr/christian.wolf/teaching



Our models are getting more complex.

How can we visualize the knowledge acquired from data?



Visualizing feature map activations



Visualizing feature map activations



Visualizing feature map activations



[Zeiler and Fergus, ECCV 2014]

Earlier layers train earlier

Evolution of features during training.



ille.	1000		0.00	C.R.B.F.	(OD) search	000	(00D)			1		20	S.	\$
10-	1	R.ª	K	*	X.	X		2	31		-	122	1228	
	- He	1	0	100	Site.	and the second			Øy	10		(m	0	۲
	1	A.	1	E.	ł		to		12		100	$\langle \mathfrak{S} \rangle$	0	3
in the second	1	18	- 5		-	-330	30		-		48		10	16
	>	*	*	La	ayer 4	×	*		141	1	La	ayer 5	.9	-31

[Zeiler and Fergus, ECCV 2014] CAM



[Zhou, Khosla, Lapedriza, Oliva, Torralba, CVPR 2016]

CAM: examples



[Zhou, Khosla, Lapedriza, Oliva, Torralba, CVPR 2016]

UIUC Event8

Guided Backpropagation

We can directly calculate the influence of each individual input pixel to a given feature map layer by calculating the gradient of this layer w.r.t. to the input signal:

$$\frac{\partial A_{i,j}^k}{\partial x_{m,n}}$$

 $A_{i,j}^k \dots$ feature cell (i,j) of a given map k. $x_{m,n} \dots$ input pixel (i,j).

Guided Backpropagation: effects on class

We can derive the output for class k w.r.t. the input pixels:

 $\frac{\partial y^c}{\partial x_{m,n}}$

 $u^c \dots$ network output for class c given input image x $x_{m,n} \dots$ input pixel (m, n).

Example: document analysis

Task: detection of text line bounding boxes. Derivatives w.r.t. to the 4 outputs (left, right, top, bottom)

Myline GONLANNES	Myline GONLAWES	Myline GONLANNES
1 Rue des Fleias	1 Rue des Flewas	1 Rue des Fleias
67250 ASCHBACH	67250 ASCHBACH	67250 ASCHBACH
Tél: D3 60536888	Tél: 03 60536888	Tél: 03 60536888
Phone line 51 Run Joo Panneso 67250 RSCHBACH	Phone line 51 Run Joo Pannes 67250 ASCHBACH	Phone line SI Run Jos Pannes G7250 RSCHBACH
Aschbach, de oz hozt zedz	Aschbach, de oz koit 2012	Aschbach, le oz Moùt Zodz
<u>Oget:</u> Fachure Réplonique avonée	Objet : Færue réléphonique enonée	Oget: Fæhre Eléphnique evenies
Madare, Monsieur, 2 constate soon me facture portant le russéro 1524817 et date du 31 juillet 2012 une concommation Villiphonique avonatement élevée. Selon le détail de cité fature, floreurs numéros énuméris me unt étrages. 2 souhaitrais den vienent qu'un devos agents me contaite des que l'on 1 souhaitrais den vienent qu'un devos agents me contaite des que l'on 1 souhaitrais den vienent qu'un devos agents me contaite des neues misse procéder à la vérificultion de me étre léléphonique et pour me missione non les provenance de numéros contenues dans no fature. 1 dons l'attact de voire étense, veullez agréer, Radame, Rensieur, me sentiments destingués. Une conceves	Madare, Marieur, Le constate sear me facture porton le nucéro 1524813 et date su 31 juliet 2012 une concommention l'éléphonique avanuelement éturée. Selon le détaite de uté facture, planeaux nuréros énanciers me contrete den quellon le subaiterais lonc vincent qu'un de vos agrito me contrete den quellon huisse procédur de venfriculton se ma conte éléphonique et pour me huisse procédur de venfriculton se ma conte éléphonique et pour me interve que mois les processors de marchos containes dans au plane. Le de content ou voire reports, vencles agrit d'adare remainer, me interventions que Mytérie CONCENTES MA	Madare, Monieur, J. constate serve me facture, porton to number 1524817 et date du 31 fuillet 20 22 une concommation l'éléphonique avanuatement élevé . Selon le détait de utte facture, fluereux numbers énancérés me cont étraiges. Je subaits aise ten vision of vien de vos agrite me contate afin quellon misse procéder à le vénfereurin some avaitement étraiges. Tense que sons le prevenance contaire de vientes rense que sons le prevenance contaire de vientes de l'alterte à voire oper vientes gour miningen à étate de l'alterte à voire oper vientes gour miningen à étate Unite contaires Unite contaires

Left

Example: document analysis



Тор

Bottom

The network looks at the right side to regress the top coordinate: it figured out line slant!

[Bastien Moysset, PhD Thesis, 2018]

Visualizing high-dimensional spaces

Problem: many tensors (input images or signals, intermediate feature maps etc.) are embedded in highdimensional spaces.

Humans cannot imagine or visualize more than 3D easily.

Can we find a mapping to a lower dimensional space which approximates the structure of the original space?

Close points should stay close.

Far points should stay far.

Example: MNIST images = 28x28 pixels = 784dim input space



t-SNE

t-distributed stochastic neighbor embedding.



t-SNE: model

Data points in input space: x_i . Data points in low-dim space: y_i .

We assign a conditional probability to the (assymetric) pair of **high-dim** datapoint (x_i, x_j) : Given x_i , will x_j be selected as neighbor if neighbors are selected according to distance (using a Gaussian kernel with variance σ_i):

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)}$$

A symmetric version is given as:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$

Van der Maaten, Hinton, JMLR 2008

t-SNE: model

For points (x_i, x_j) in the **lows-dim** map, we use a Cauchy distribution, which is heavier tailed:

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}}$$

Reason:

- high-dim spaces are less crowded (the volume of a sphere of radius r and dim d grows with r^d).
- Heaver tails better model datapoints which are not too close away from each other.

Van der Maaten, Hinton, JMLR 2008

t-SNE: training

We solve for the points y_i , $\forall i$ using SGD and optimizing Kullback-Leibler divergence (KL):

$$C = KL(P||Q) = \sum_{i} \sum_{j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Van der Maaten, Hinton, JMLR 2008

t-SNE: Hyper-parameter perplexity

We need to set the σ_i parameter for the distribution P(i) each datapoint i.

A global hyper-parameter Perp (=*Perplexity*) is set by the user (~ number of neighbors of each point).

Then se solve for

$$\operatorname{Perp}\left(P_{i}\right) = 2^{H(P_{i})}$$

where

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}$$

Example: the breakfast action data



4 labels per video clip:

- The recipe (e.g. cesar salade)
- The short term action (e.g. cut chicken)
- The person performing the action
- The camera viewpoint

[H. Kuehne, A. B. Arslan and T. Serre. The Language of Actions: Recovering the Syntax and Semantics of Goal-Directed Human Activities. CVPR, 2014.]

t-SNE: activity recognition

Non-fine tuned on train datase



Work of Tom Gillooly Breakfast Dataset (train split), before fine-tuning

t-SNE: activity recognition

Fine tuned on train datase



Work of Tom Gillooly Breakfast Dataset (train split), after fine-tuning