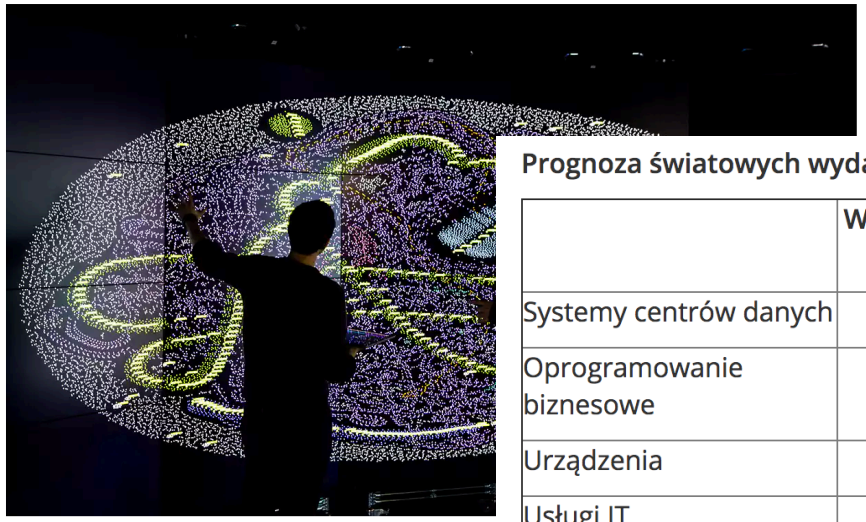
Siamese architectures applied to semantic sentence similarity comparison tasks

Michał Sitko

Big Picture

Scientists plan huge European AI hub to compete with US

Exclusive: In an open letter, the scientists say the proposed Ellis institute is essential to avoid brain drain to big tech firms



▲ Scientists inspecting AI work at the Data Science Institute at Imp
for the Guardian

European Commission Calls for €20bn Cash Injection to Push AI Research

The European Commission has called for a cash injection in AI funding to stop it falling behind the US and China in AI innovation.

Prognoza światowych wydatków na IT w latach 2017 - 2018

	Wydatki w 2017 r. (mld dol.)	Wzrost w 2017 r. (%)	Wydatki w 2018 r. (mld dol.)	Wzrost w 2018r (%)
Systemy centrów danych	173	1,7	176	1,8
Oprogramowanie biznesowe	354	8,5	387	9,4
Urządzenia	664	5,3	697	5,0
Usługi IT	931	4,0	980	5,3
Usługi telekomunikacyjne	1387	0,9	1417	2,2
Łącznie IT	**3508**	**3,3**	**3658**	**4,3**

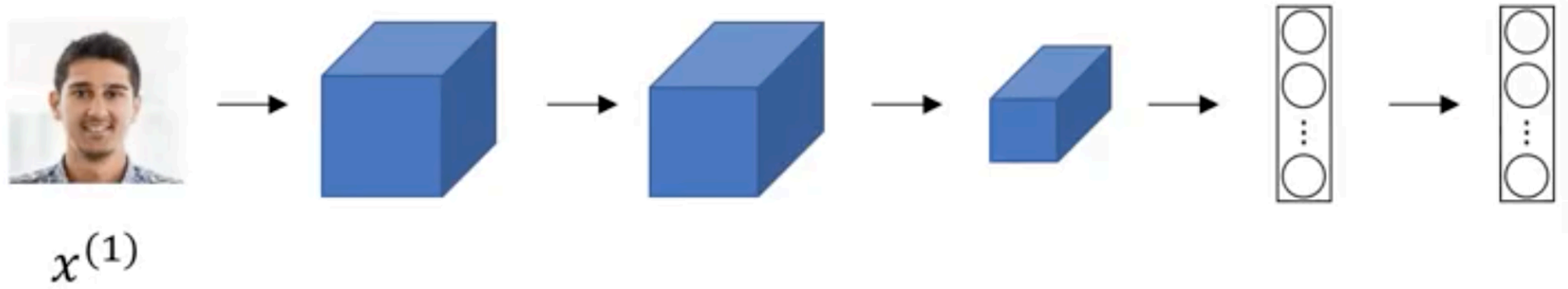
Źródło: Gartner

Agenda

- Siamese models
  - Applications
- Memory models
  - Applications
- Word embeddings
- SemEval STS Excercise

Siamese models

Siamese network



Signature Verification – J. Bromley et al. 1994

feature 1 pen up = −1 ; pen down = +1, (pud)

feature 2 x position, as a difference from the linear estimate for $x(t)$, normalized using the standard deviation of y, (x)

feature 3 y position, as a difference from the linear estimate for $y(t)$, normalized using the standard deviation of y, (y)

feature 4 speed at each point, (spd)

feature 5 centripetal acceleration, (acc-c)

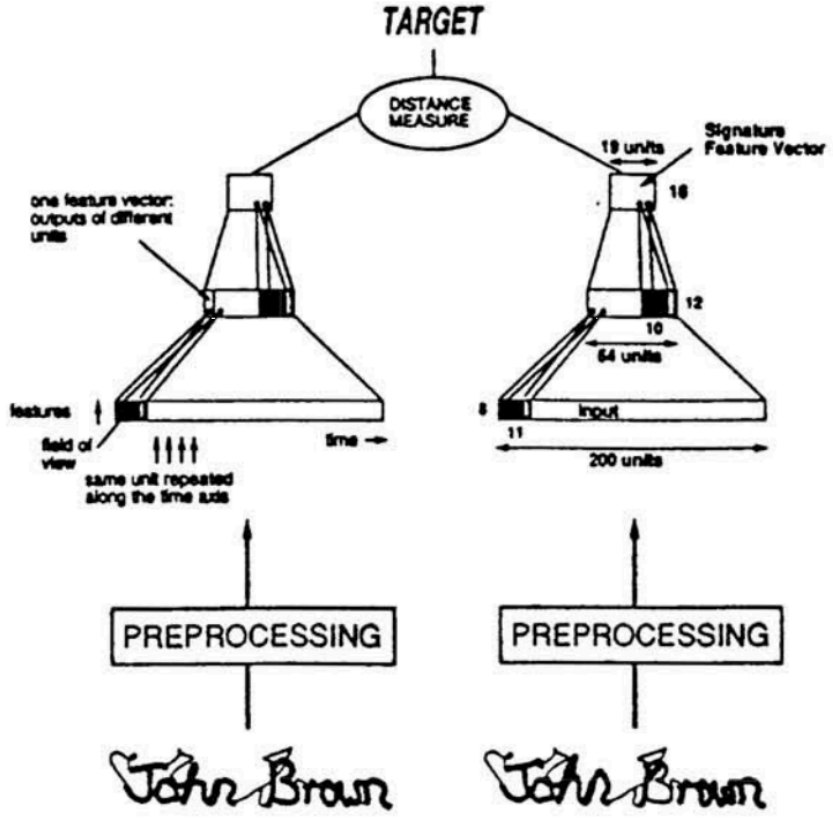
feature 6 tangential acceleration, (acc-t)

feature 7 the direction cosine of the tangent to the trajectory at each point, ($\cos\theta$)

feature 8 the direction sine of the tangent to the trajectory at each point, ($\sin\theta$)

feature 9 cosine of the local curvature of the trajectory at each point, ($\cos\phi$)

feature 10 sine of the local curvature of the trajectory at each point, ($\sin\phi$)

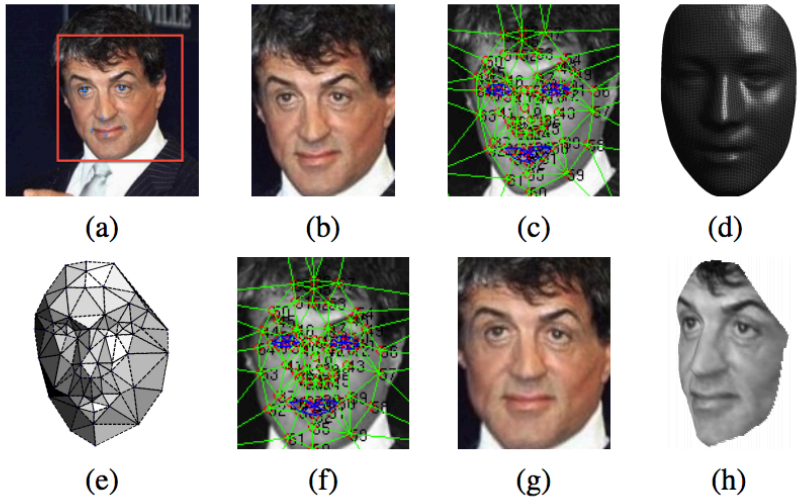


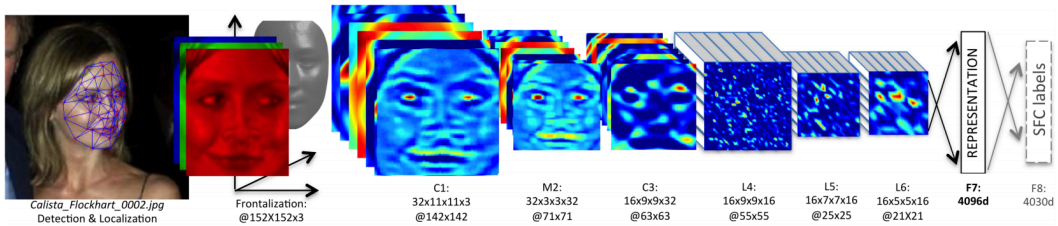
Signature Verification – J. Bromley et al. 1994

Training

- Loss function = Similarity (dissimilarity) function:
  - cosine distance,
  - manhattan distance,
  - cross-entropy loss (?),
  - contrastive loss, http://yann.lecun.com/exdb/publis/pdf/hadsell-chopra-lecun-06.pdf
  - center loss. https://medium.com/mlreview/experiments-with-a-new-loss-term-added-to-the-standard-cross-entropy-85b080c42446
- Split the loss over neural network branches and propagate using backpropagation algorithm

DeepFace – Facebook – 2014

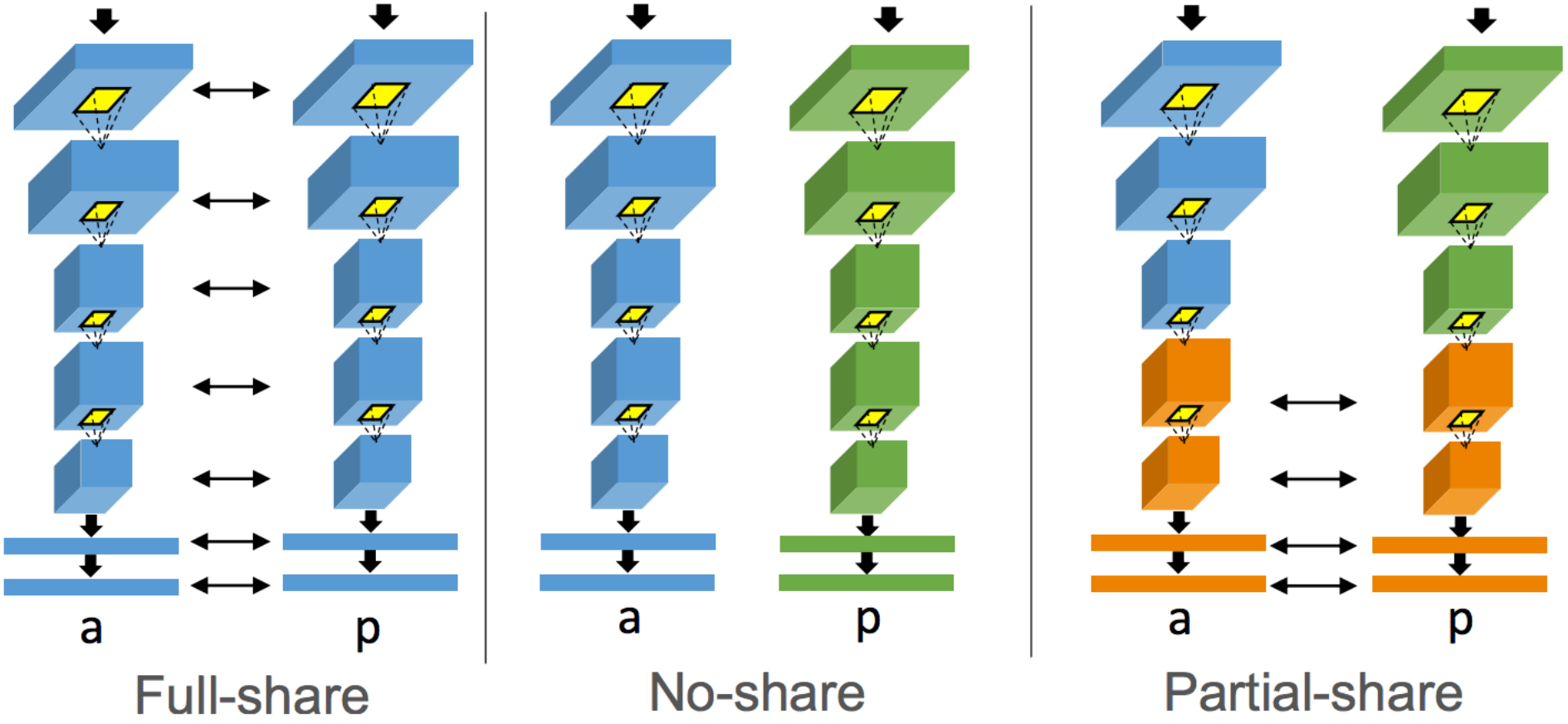




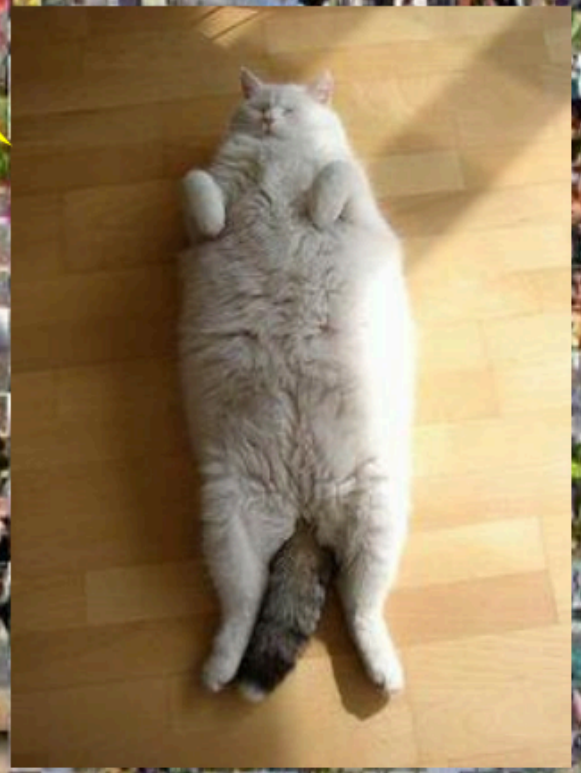
Network	Error (*SFC*)	Accuracy ± SE (*LFW*)
DeepFace-align2D	9.5%	0.9430 ±0.0043
DeepFace-gradient	8.9%	0.9582 ±0.0037
DeepFace-Siamese	NA	0.9617 ±0.0038

Method	Accuracy ± SE	Protocol
Joint Bayesian [6]	0.9242 ±0.0108	restricted
Tom-vs-Pete [4]	0.9330 ±0.0128	restricted
High-dim LBP [7]	0.9517 ±0.0113	restricted
TL Joint Bayesian [5]	0.9633 ±0.0108	restricted
DeepFace-single	**0.9592** ±0.0029	unsupervised
DeepFace-single	**0.9700** ±0.0028	restricted
DeepFace-ensemble	**0.9715** ±0.0027	restricted
DeepFace-ensemble	**0.9735** ±0.0025	unrestricted
Human, cropped	0.9753	

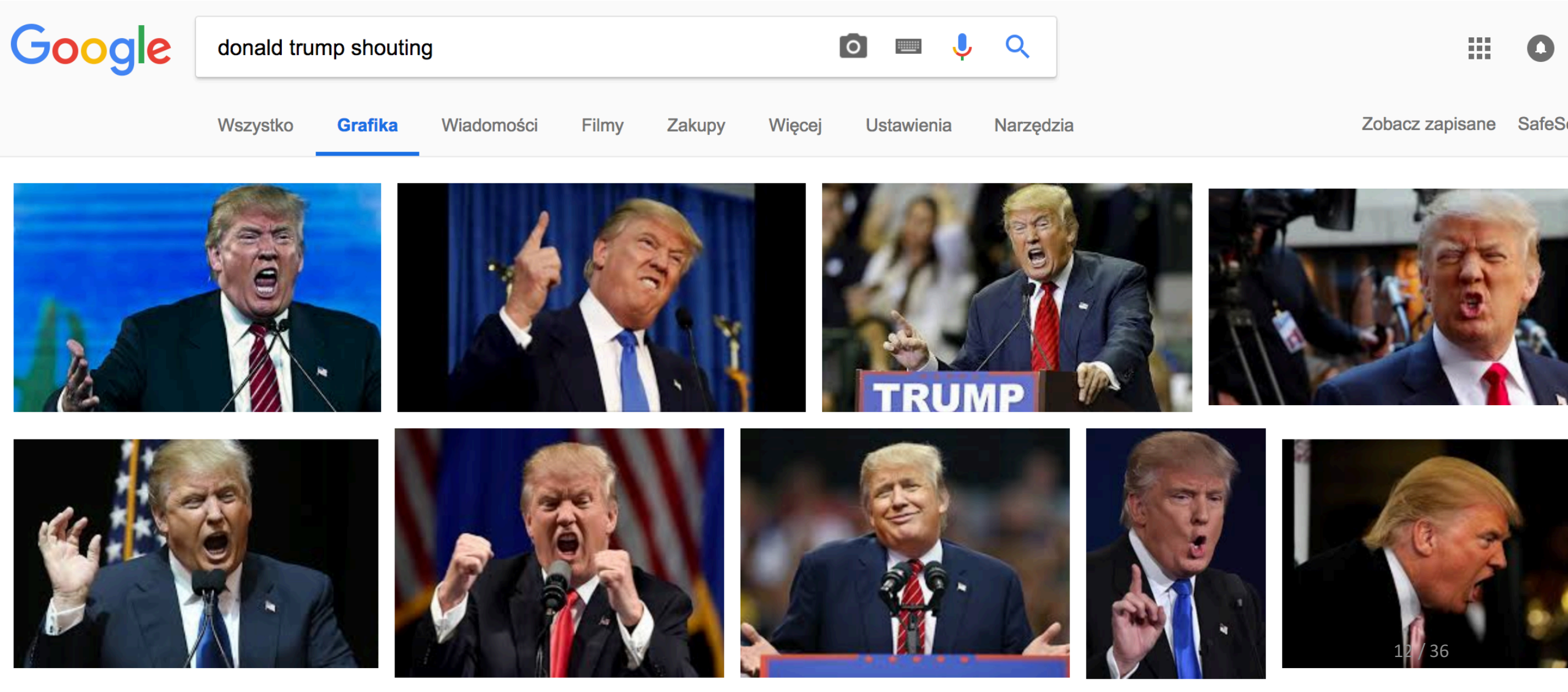
Weights sharing approaches



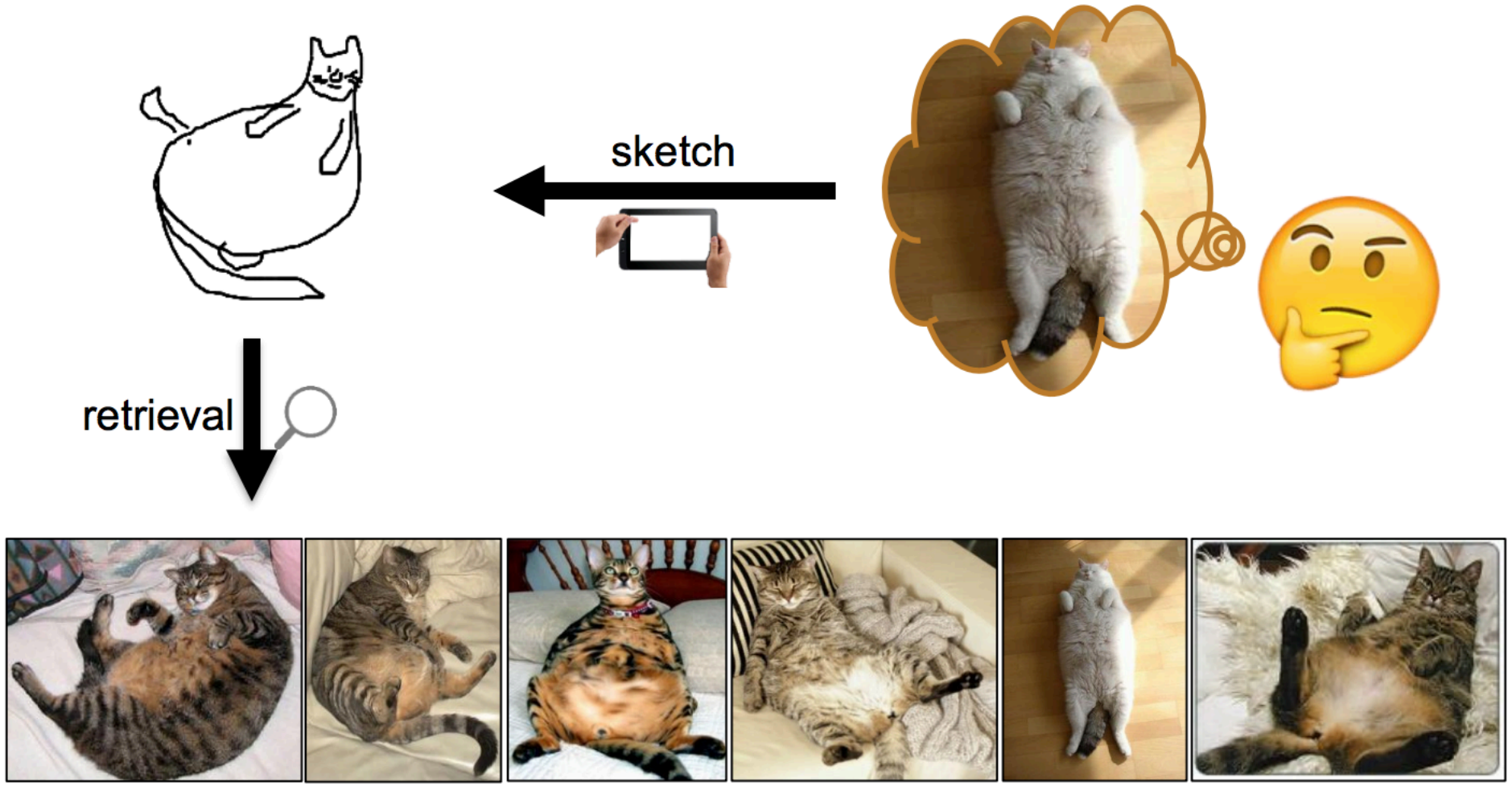
Search for a particular image in your mind?



Text Search

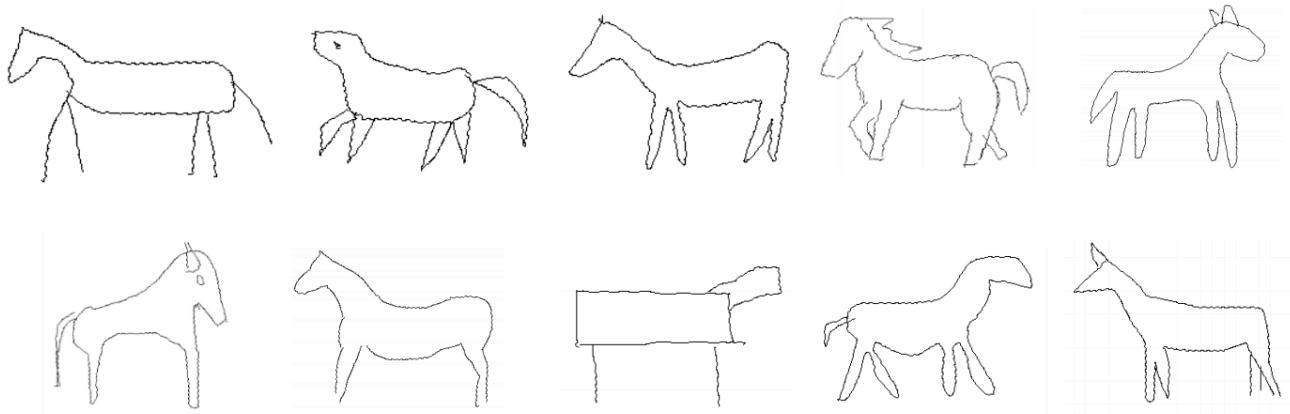

Google
donald trump shouting
Wszystko
Grafika
Wiadomości
Filmy
Zakupy
Więcej
Ustawienia
Narzędzia
Zobacz zapisane
TRUMP

Sketch-based image retrival (SBIR)

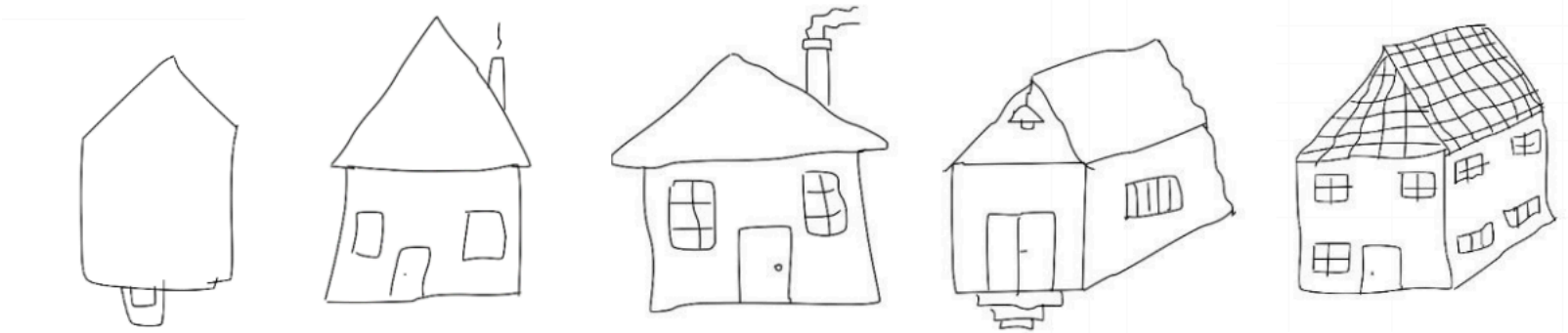


Challanges

- sketches are usually messy and inaccurate,
- sketches can have various levels of abstraction,
- does not represent real-life objects.

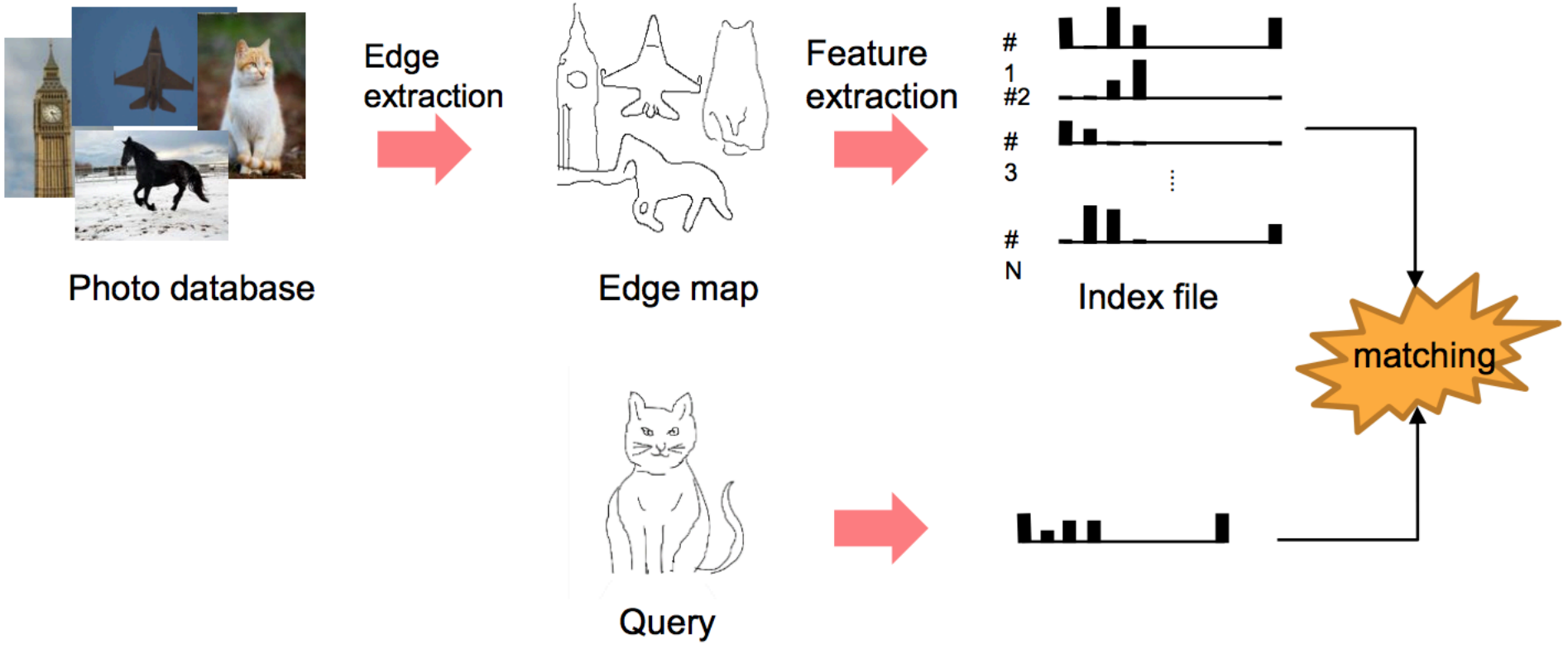


Horse

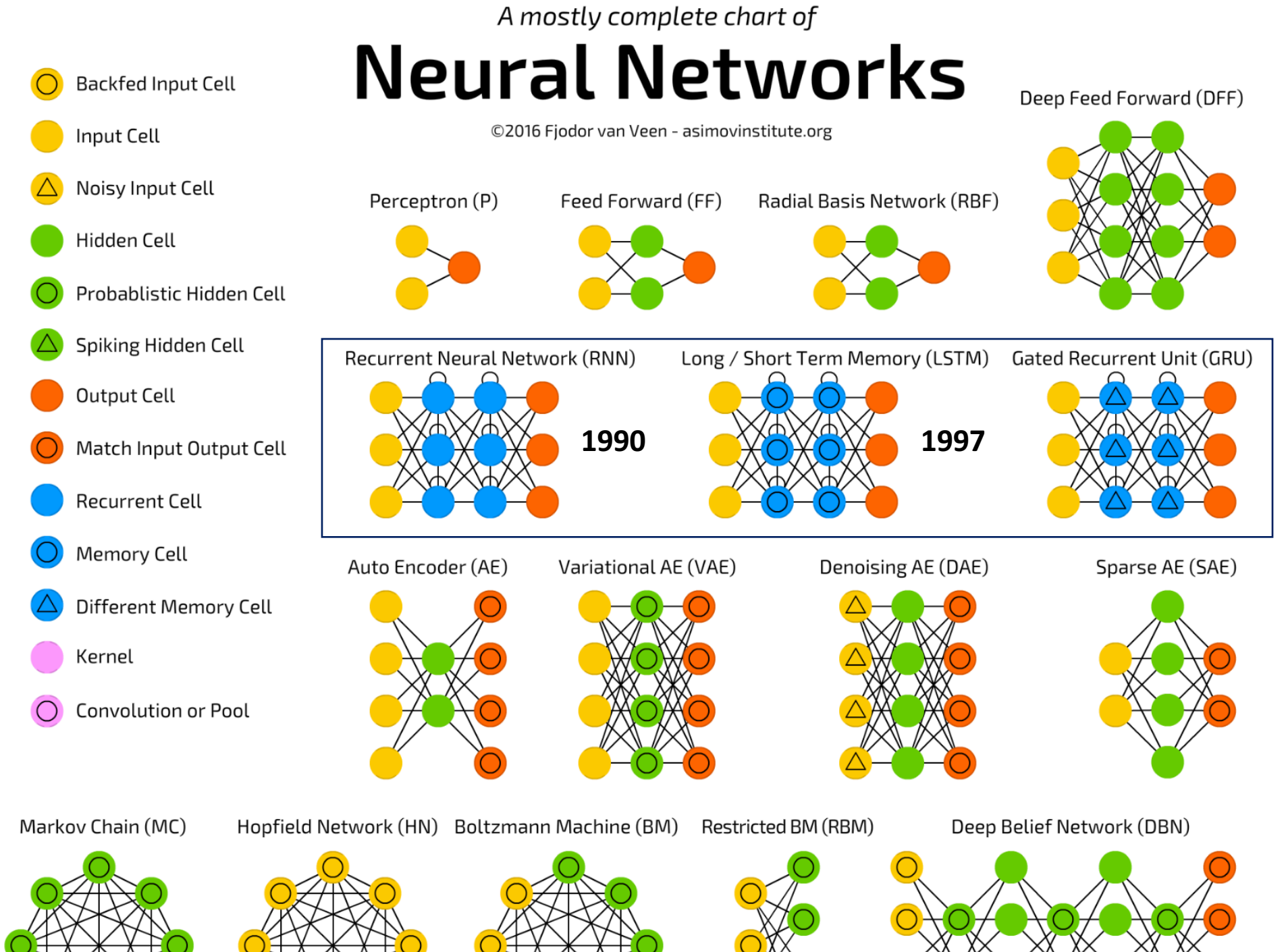


House

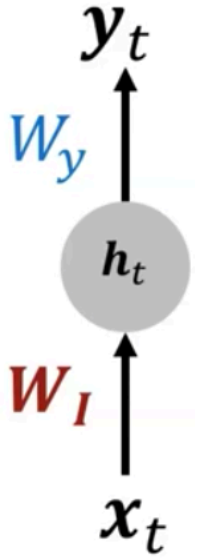
Traditional approach



Memory models

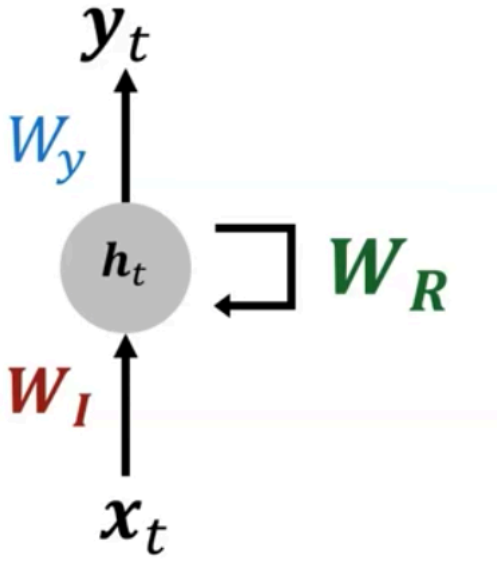

A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
1990
Long / Short Term Memory (LSTM)
1997
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)

Classic neuron



$$\boldsymbol{y}^{(t)} = g_y(W_{\mathrm{I}}\boldsymbol{x}^{(t)} + \boldsymbol{b}_y)$$

Recurrent neuron



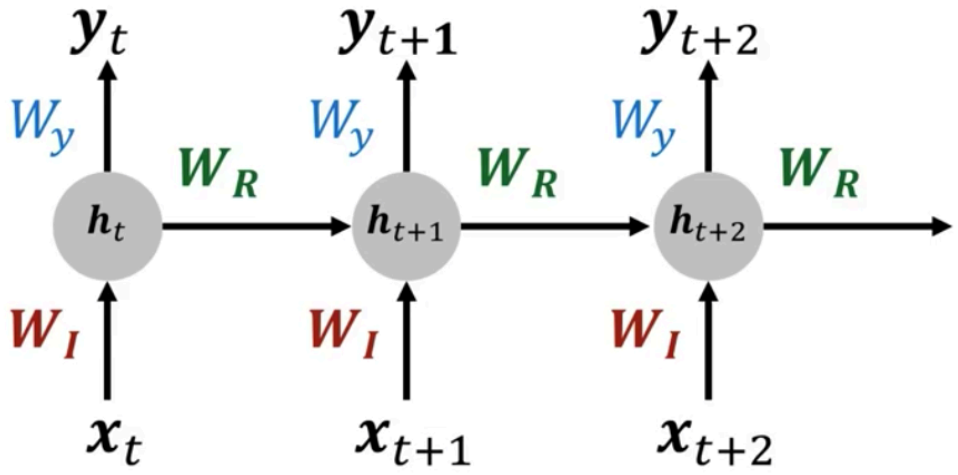
$$\boldsymbol{h}^{(t)} = g_h(W_{\mathrm{I}}\boldsymbol{x}^{(t)} + W_{\mathrm{R}}\boldsymbol{h}^{(t-1)} + \boldsymbol{b}_h)$$

$$\boldsymbol{y}^{(t)} = g_y(W_y\boldsymbol{h}^{(t)} + \boldsymbol{b}_y)$$

Unrolling recurrent network

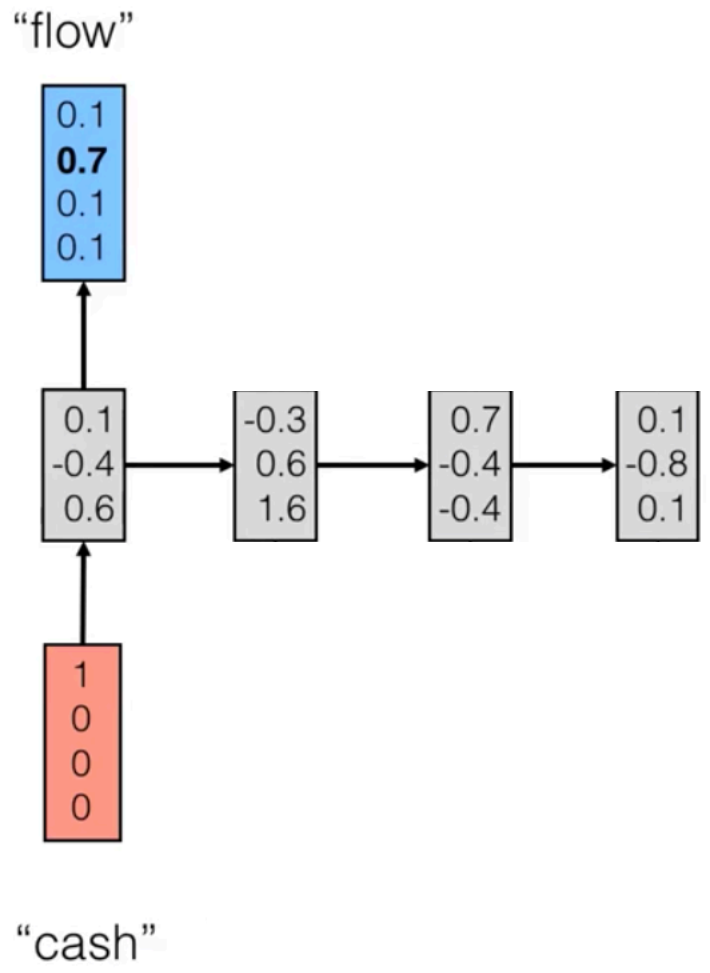
$$\boldsymbol{h}^{(t)} = g_h(W_{\mathrm{I}}\boldsymbol{x}^{(t)} + W_{\mathrm{R}}\boldsymbol{h}^{(t-1)} + \boldsymbol{b}_h)$$

$$\boldsymbol{y}^{(t)} = g_y(W_y\boldsymbol{h}^{(t)} + \boldsymbol{b}_y)$$



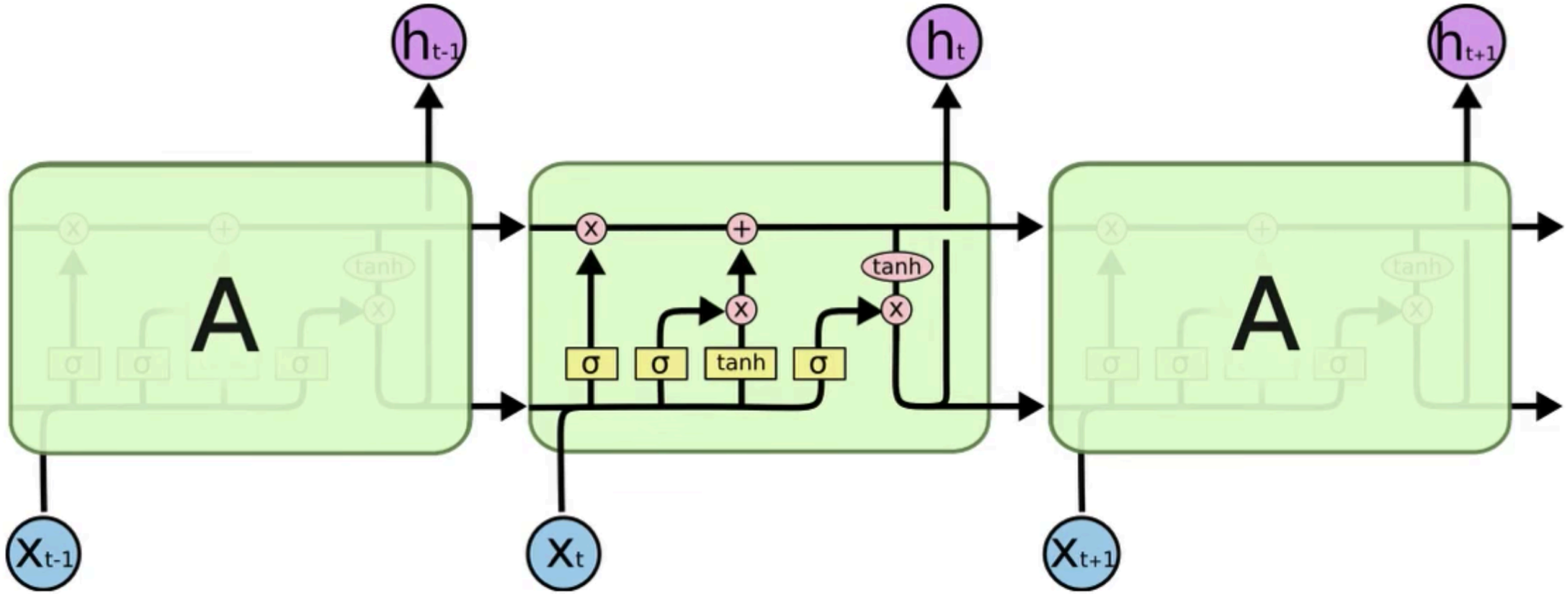
Applications

- Learning language models,
- Sentiment analysis,
- Machine translation.



Recurrent Neural Networks, Intel Nervana, https://www.youtube.com/watch?v=Ukgii7Yd_cU

LSTM



http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Examples:

Shakespeare

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Examples

LaTeX

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma **??**. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

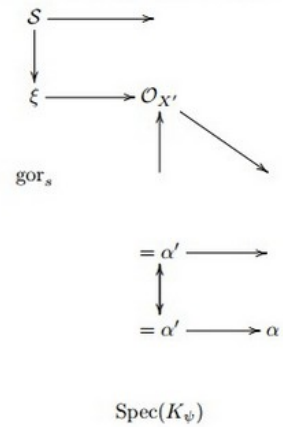
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas **??**.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition **??** and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

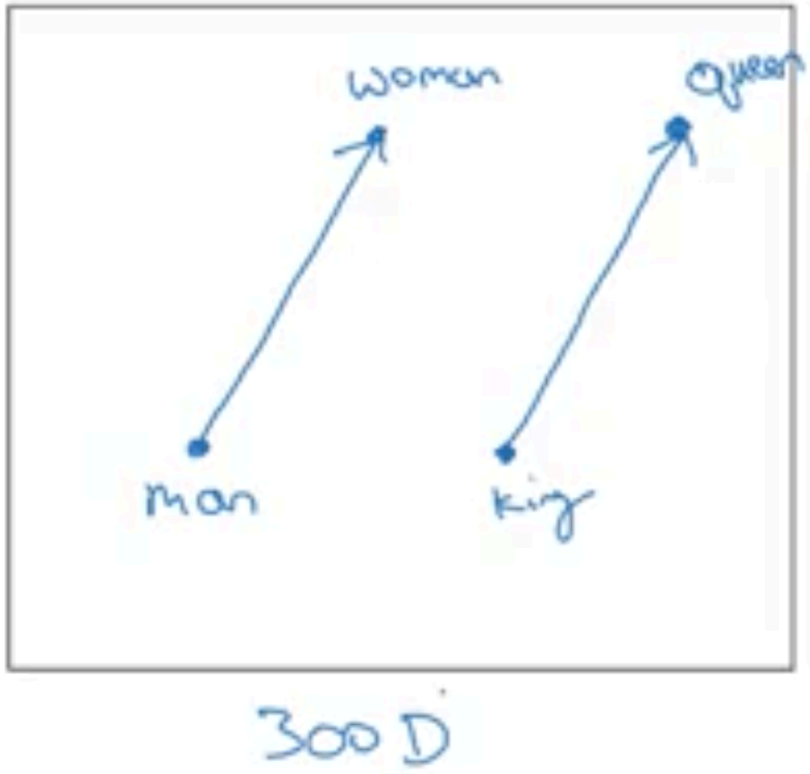
If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma **??**. This is a sequence of $\mathcal{F}$ is a similar morphism.

Word Embeddings

Word Embeddings

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	−1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97

Word Embeddings



$$e_{man} - e_{woman} \approx e_{king} - e_?$$

SemEval STS Exercise

Structural & Semantic similarity

- He is smart = He is a wise man.
- Someone is travelling countryside = He is travelling to a village.
- She is cooking a dessert = Pudding is being cooked.
- Microsoft to acquire Linkedin ≠ Linkedin to acquire microsoft

Gold Standard

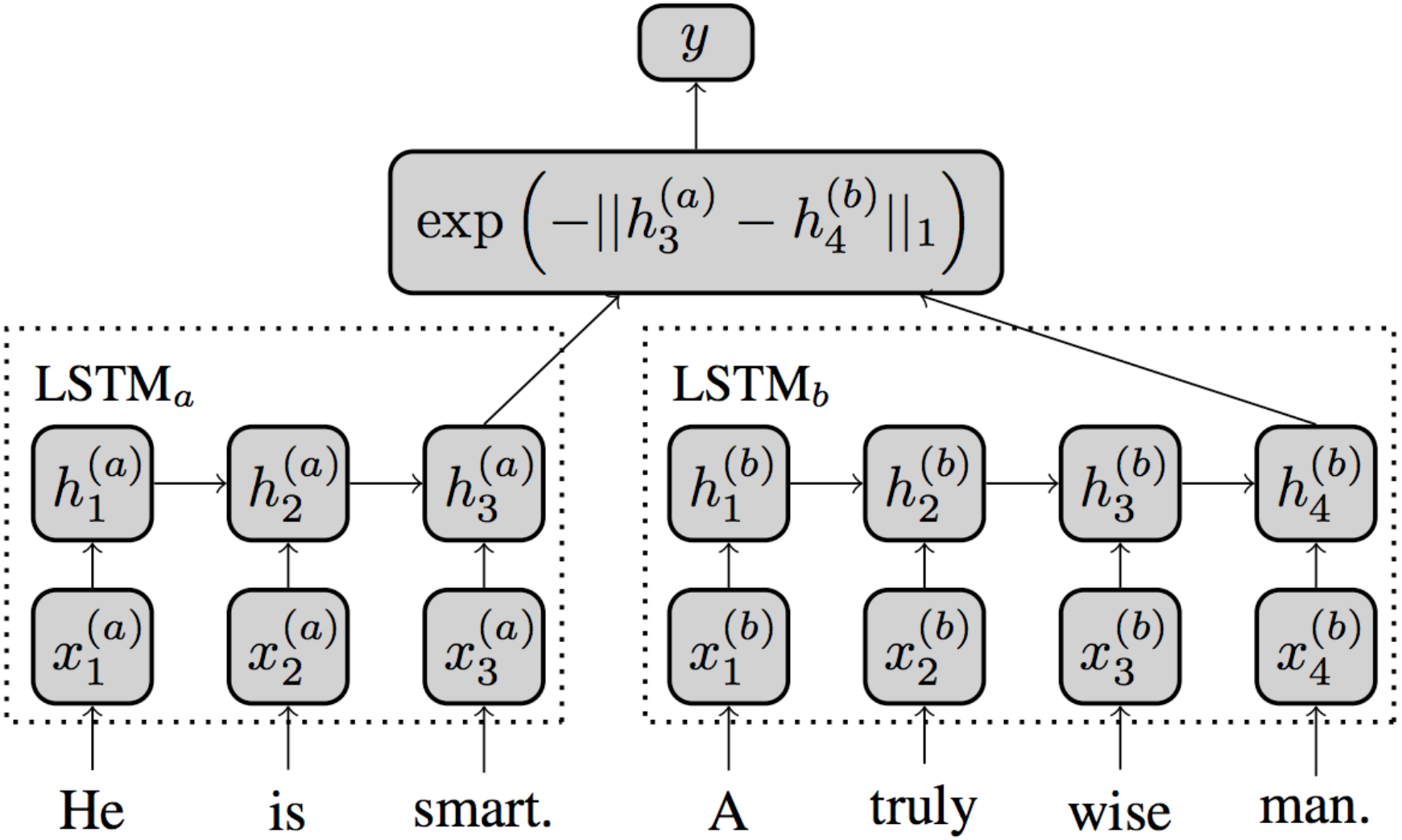
- (5) The two sentences are completely equivalent, as they mean the same thing.
  The bird is bathing in the sink.
  Birdie is washing itself in the water basin.
- (4) The two sentences are mostly equivalent, but some unimportant details differ.
  In May 2010, the troops attempted to invade Kabul.
  The US army invaded Kabul on May 7th last year, 2010.
- (3) The two sentences are roughly equivalent, but some important information differs/missing.
  John said he is considered a witness but not a suspect.
  "He is not a suspect anymore." John said.
- (2) The two sentences are not equivalent, but share some details.
  They flew out of the nest in groups.
  They flew into the nest together.
- (1) The two sentences are not equivalent, but are on the same topic.
  The woman is playing the violin.
  The young lady enjoys listening to the guitar.
- (0) The two sentences are on different topics.
  John went horse back riding at dawn with a whole group of friends.
  Sunrise at dawn is a magnificent view to take in if you wake up early enough for it.

STS 2017 en-en

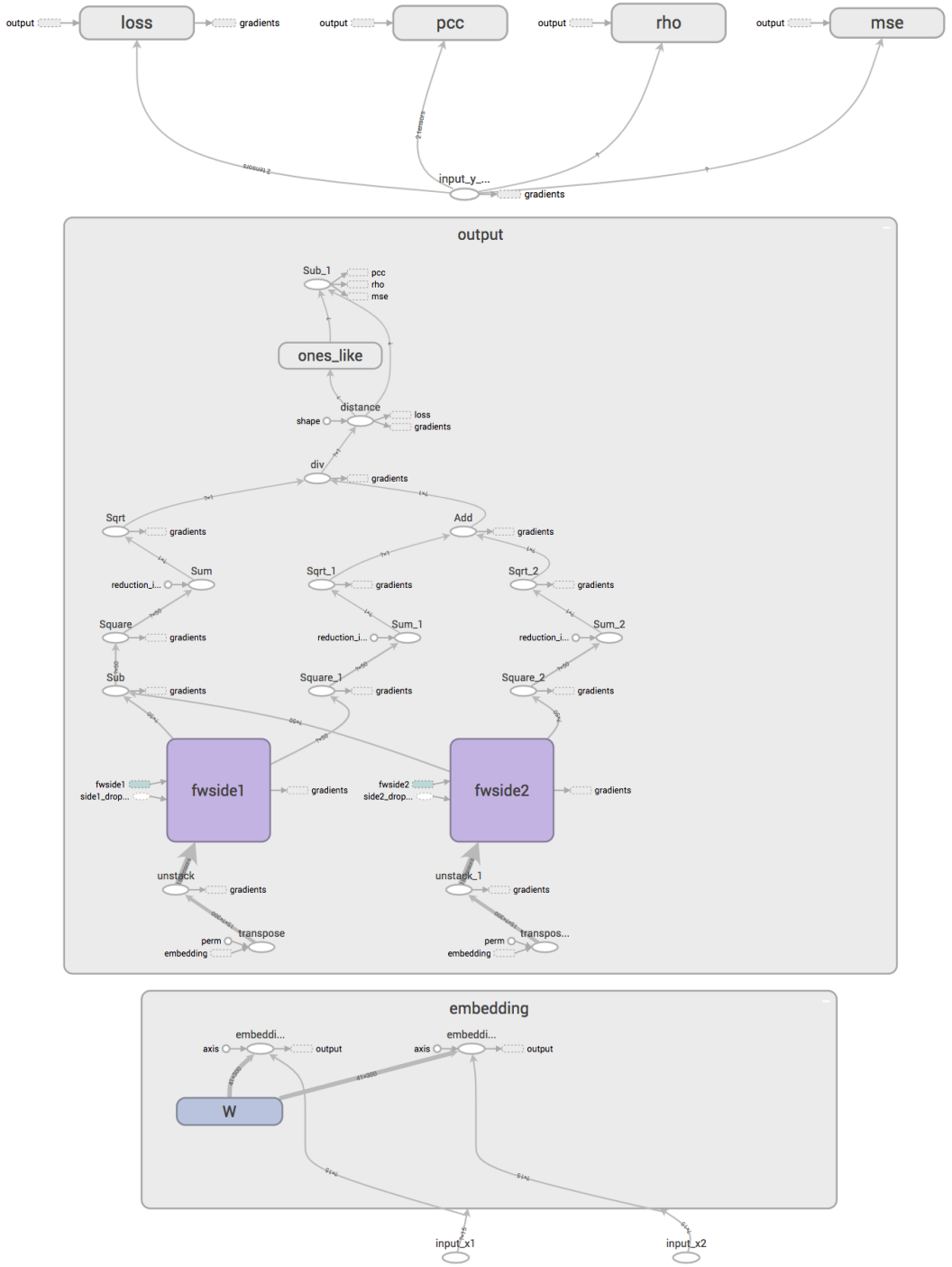
```
2.400000    A person is on a baseball team. A person is playing basketball on a team.
0.200000    Our current vehicles will be in museums when everyone has their own aircraft.   The car needs to some work
1.000000    A woman supervisor is instructing the male workers. A woman is working as a nurse.
2.000000    A bike is next to a couple women.   A child next to a bike.
2.200000    The group is eating while taking in a breathtaking view.    A group of people take a look at an unusual tree.
3.400000    The boy is raising his hand.    The man is raising his hand.
0.800000    A man with a gray beard is being shaved in front of a lecture hall  A man with a beard is sitting in the grass.
0.400000    The sky has very little to no clouds.   This Lady might be ready for Rock Climbing, or just watching the Clouds, above.
0.400000    The young boy jumps barefoot outside in the front yard. The teen rode his bike around the people walking in the courtyard.
4.000000    There are dogs in the forest.   The dogs are alone in the forest.
3.400000    Some cyclists stop near a sign. Two men stop to talk near a sign outside.
3.000000    These cooks in the white are busy in the kitchen making dinner for their customers. The women are preparing dinner in their
4.600000    A young person  deep in thought.    A young man deep in thought.
```

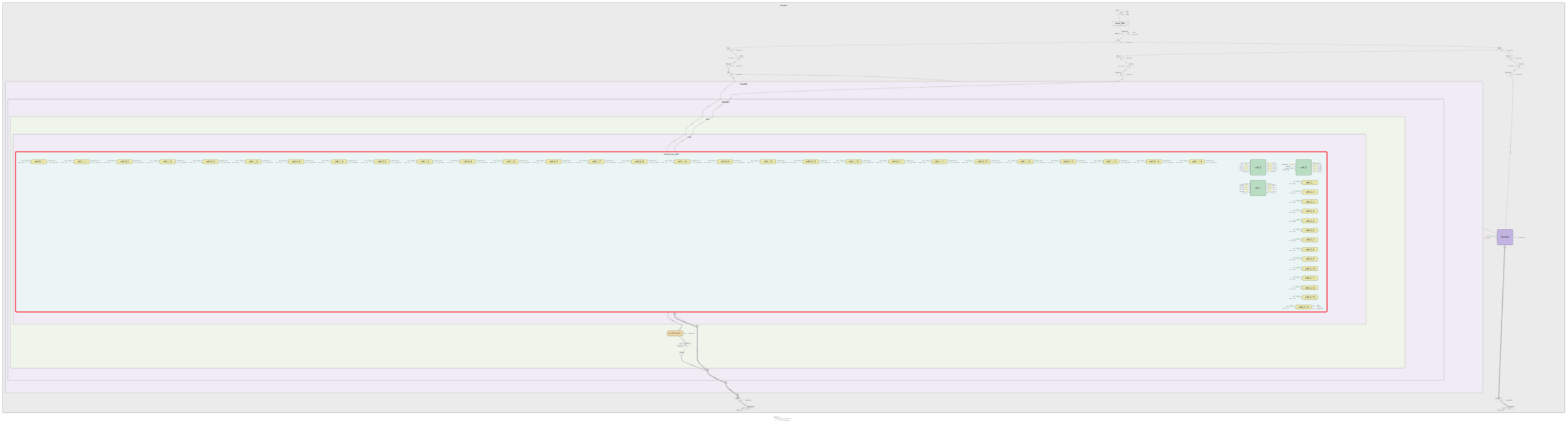
STS 2017 en-ar

1.000000	A tired businessman takes a quick nap while riding the train.	يجد رجل الوقت لأخذ غفوة في سيارته الكهربائية.
0.200000	A man working on a light that has been broken for a week.	كان يوم عمل المرأة شاقا وهي ملتوية للأسفل .
0.400000	A man is attempting to fix the air compressor.	رجل يستخدم خرطوما فائق القوة لإخماد الحريق .
4.200000	a twelve year old girl eats chocolate ice cream	فتاة تأكل آيس كريم.
3.400000	a man holds a pro Judea protest sign at a protest rally	رجل يحمل لافتة احتجاج
4.400000	A woman holds a little boy wearing a blue sweatshirt.	امرأة تحمل طفلا مبتسما يرتدي قميصا أزرق.
3.200000	A bike riding blond child.	الفتاة تركب دراجتها .
2.600000	people are in the road	الناس بالقرب من الطريق
1.600000	The windows are tall.	النوافذ مفتوحة.
3.000000	A man is being chased by the police.	تعتقل الشرطة رجلا
1.600000	A girl has binoculars	فتاة شابة تحمل كاميرا
1.400000	Two people ride red bikes.	رجلان يقودان إلى آخر الطريق
1.200000	Two people are standing next to each other wearing pink.	هناك ثلاثة رجال، اثنان منهما يرتديان قميصين أحمرين يثرثران معا
1.400000	A black woman is buying donuts at the store.	المرأة تنظر إلى الشامبو في محل للبقالة.
3.000000	The man works in the local market.	مجموعة من التجار يعملون في السوق.
3.600000	A woman in red wears glasses and long earrings.	امرأة تلبس نظارات، وأقراطا كبيرة، وغطاء على رأسها.
4.000000	Skateboarder in midair at a park.	رجل في الهواء الطلق حديقة تزلج.



Siamese Recurrent Architectures for Learning Sentence Similarity, Jonas Muller et al. 2016


output
loss
gradients
output
pcc
output
rho
output
mse
input_y_...
gradients
output
Sub_1
pcc
rho
mse
ones_like
distance
shape
loss
gradients
div
gradients
Sqrt
gradients
Add
gradients
Sum
reduction_i...
Sqrt_1
gradients
Sqrt_2
gradients
Square
gradients
Sum_1
reduction_i...
Sum_2
reduction_i...
Sub
gradients
Square_1
gradients
Square_2
gradients
fwside1
side1_drop...
fwside1
gradients
fwside2
side2_drop...
fwside2
gradients
unstack
gradients
unstack_1
gradients
perm
embedding
transpose
perm
embedding
transpos...
embedding
embeddi...
axis
output
embeddi...
axis
output
W
input_x1
input_x2



Results

Method	r	ρ	MSE
Illinois-LH (Lai and Hockenmaier 2014)	0.7993	0.7538	0.3692
UNAL-NLP (Jimenez et al. 2014)	0.8070	0.7489	0.3550
Meaning Factory (Bjerva et al. 2014)	0.8268	0.7721	0.3224
ECNU (Zhao, Zhu, and Lan 2014)	0.8414	–	–
Skip-thought+COCO (Kiros et al. 2015)	0.8655	0.7995	0.2561
Dependency Tree-LSTM (Tai, Socher, and Manning 2015)	0.8676	0.8083	0.2532
ConvNet (He, Gimpel, and Lin 2015)	0.8686	0.8047	0.2606
MaLSTM	**0.8822**	**0.8345**	**0.2286**

Siamese Recurrent Architectures for Learning Sentence Similarity, Jonas Muller et al. 2016

References

- **Signature Verification using a "Siamese" Time Delay Neural Network,** J. Bromley et. al. 1994
  https://papers.nips.cc/paper/769-signature-verification-using-a-siamese-time-delay-neural-network.pdf
- **DeepFace: Closing the Gap to Human-Level Performance in Face Verification,** Y. Taigman et. al 2014
  https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf
- **Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask**
  http://conteudo.icmc.usp.br/pessoas/moacir/p17sibgrapi-tutorial/#CODE
- **Siamese architecture, Tomasz Trzciński @ PL in ML conference 2017**
  https://www.youtube.com/watch?v=xe-NBP_zUMM&t=2s
- **Recurrent Neural Networks, Intel Nervana,**
  https://www.youtube.com/watch?v=Ukgii7Yd_cU
- **The Unreasonable Effectiveness of Recurrent Neural Networks, Andrej Karpathy 2015**
  http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- **Understanding LSTM, Cristopher Olah**
  http://colah.github.io/posts/2015-08-Understanding-LSTMs/
- **RNN15. NLP Properties of word embeddings, deeplearning.io Andrew Ng 2018**
  https://www.youtube.com/watch?v=fmfpwyyg6mQ
- **Siamese Recurrent Architectures for Learning Sentence Similarity, Jonas Muller et al. 2016**
  http://www.mit.edu/~jonasm/info/MuellerThyagarajan_AAAI16.pdf
- **Repository**
  https://github.com/zacateras/deep-siamese-text-similarity