

A SIMPLE AND FAST WORD SPOTTING METHOD

Prof. Lior Wolf
The School of Computer Science
Tel Aviv University

Joint work with Alon Kovalchuk and Nachum Dershowitz

Images → Text

OCR is very hard for handwritten material

Instead:

- Approximate match
- **Word spotting**
- Transcript alignment

אָזְבַּעַיִשׁ עַדְיוֹד אֲרוֹלֵי הָנָנוֹנִיך
יְדָרָא נְוִוִּיבִיס יְאָרָרָא נְחִיבָּלִי
שְׂתֻמְצְרִירָשׁ אֶלְצָאָנוֹנִיצְדוֹגָזָןָם
צָאוּבְשָׁמִידָו אֲרָרָא נְאָנָא יְזָהָלָטָס
שְׁינְוֹנִיזְוֹצְדִּיקָו אֲדִינוֹסְרָחָשָׁו
עֲרָתָאִיסְזּוֹזְלִילִילִילִידָו שְׁיָעָשָׁו
בִּינְשִׁלְסָנוֹחִיכִיכִידְגָּמְלָעִיכִי
כִּיחְלָצְתָנְשִׁיםְמוֹעָעָנִינִיסְדָּסָע
אַרְגָּלִיאַעֲדָלְלִנִּיְזָדְבָּאָרְדָּחָי
פְּלָאָסְנָעִיכִיאַדְבָּרָאָנִיגָּלִיאַעֲדָל

אהבתִי כִּי יִשְׁמַע יְהוָה אֶת
קוֹלִי תְּחִנּוּנִי כִּי הָטה אֶזְעָנוּ לִ
וּבִימִי אֲקָרָא אֲפִפּוֹנִי חָבֵל מֹות
וּמְצָרִי שְׁאָוֹל מְצָאָוֹן צְרָח
וַיְגַן אַמְצָא וּבְשָׁמֶן יְהָזָה
אֲקָרָא אָנָא יְהָזָה מְלָטָס
נְפִשְׁיָהָנָנִי יְהָזָה וְצָדִיקָא נְלָהָנִי
מְרַחַט שְׁוּמָר פְּתָאִים יְהָזָה
דְּלָוָתִילִי יְהָשָׁעָנִ שְׁבִי נְפִשְׁיָה
לְמְנָהִיכִי כִּי יְהָזָה גָּמָל עַלְיכִי
כִּי חָלָצְתָנְשִׁמְשִׁי מֹות אֶת עַנִּי
מְדֻמְעָה אֶת-רְגָלִי מִזְמָרָה
אֲתָהָלָךְ לְפָנֵי יְהָזָה בְּאֶרְצָוֹתִי
הָחַיִם הָאָמָנָתִי כִּיאֲדָבָרִ אֲנִי



OCR

Transcription

Word spotting

Given an image of a query word,
identify all other occurrences of the
same word in a set of input images

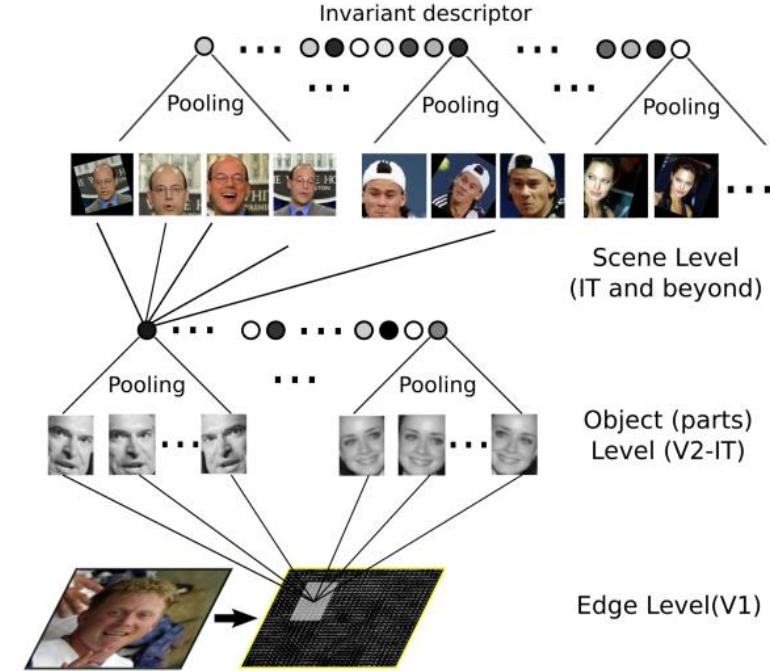
Query: Socrates

Socrates was a Classical Greek philosopher. Credited as one of the founders of Western philosophy, he is an enigmatic figure known only through the classical accounts of his students. Plato's dialogues are the most comprehensive accounts of Socrates to survive from antiquity. Forming an accurate picture of the historical Socrates and his philosophical viewpoints is problematic at best. This issue is known as the Socratic problem. The knowledge of the man, his life, and his philosophy is based on writings by his students and contemporaries. Foremost among them is Plato; however, works by Xenophon, Aristotle, and Aristophanes also provide important insights. The difficulty of finding the real Socrates arises because these works are often philosophical or dramatic texts rather than straightforward histories. Aside from Thucydides who makes no mention of Socrates or philosophers in general, there is in fact no such thing as a straightforward history contemporary with Socrates that dealt with his own time and place.

III Magic faces

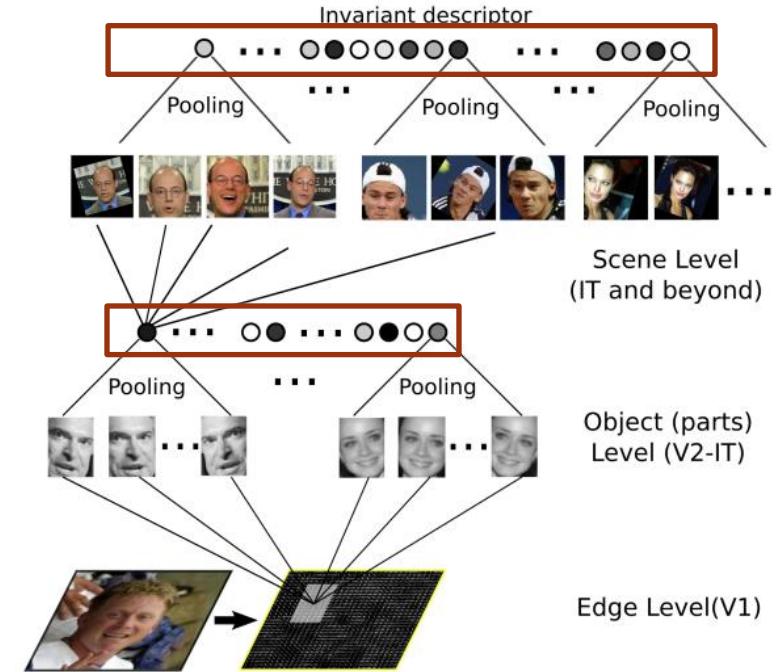
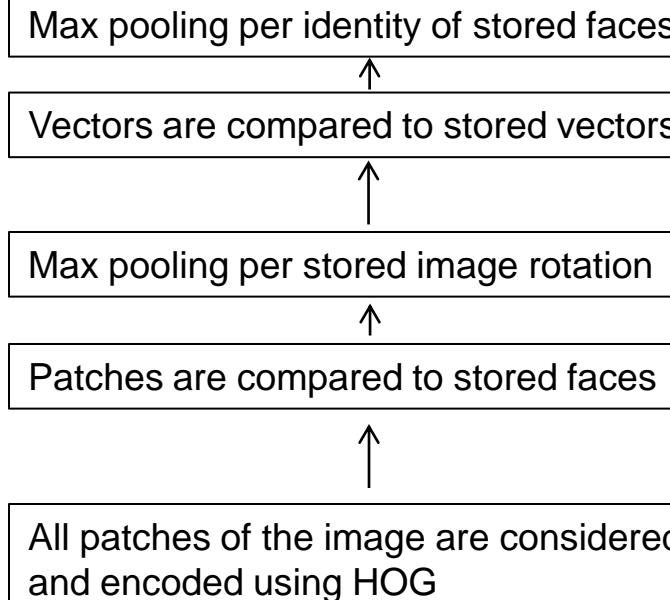
Our system resembles the biologically plausible face-recognition architecture by Poggio

That system has a very simple architecture; however, it achieves excellent performance



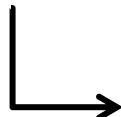
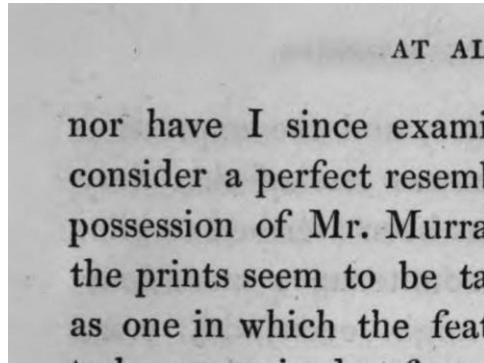
Q. Liao Q, J.Z. Leibo, Y. Mroueh, T. Poggio. **Can a biologically-plausible hierarchy effectively replace face detection, alignment, and recognition pipelines?** arXiv 2014.

III Magic faces

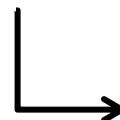
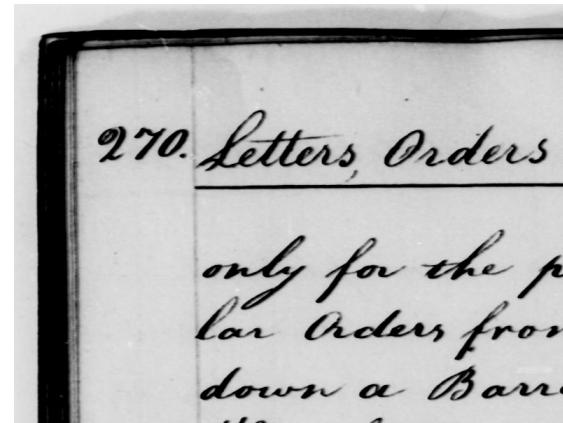


III Binarization

Simple threshold at 85% of mean pixel intensity



AT AL
nor have I since exami
consider a perfect resembl
possession of Mr. Murra
the prints seem to be ta
as one in which the feat



270. Letters, Orders

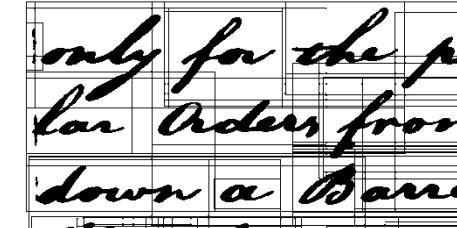
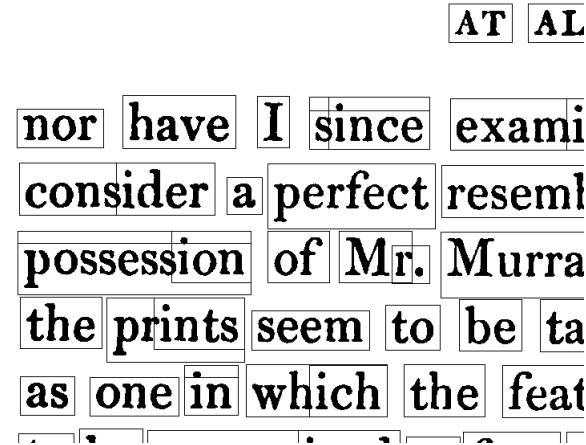
only for the p
lar Orders from
down a Barre

III Candidate patch extraction

An elaborate heuristic based on the connected components extracted from the binary image.

Connected components are grouped to form clusters that are:

- “not too big”
- continuous in the x-axis
- components are aligned in y-axis
- contains all components in its bounding box



1:2

1:30

III Patch normalization

A margin of 8 pixels is placed around the bounding box of each patch

The enlarged patches are then resized to a fixed size of 160×56 pixels

Grid of 20×7 cells, each 8×8 pixels



HOG + LBP encoding

Each cell is represented by:

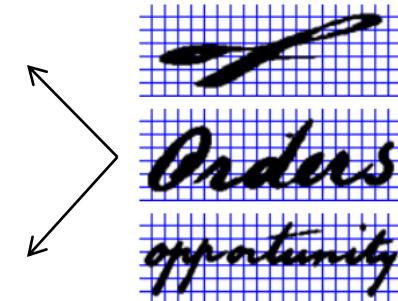
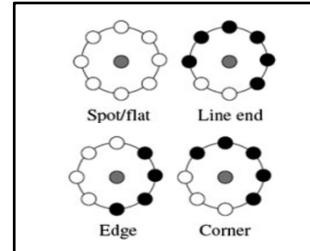
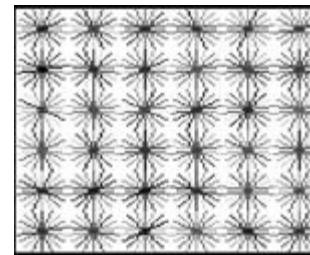
A HOG vector in R^{31}

An LBP vector in R^{58}

All HOG vectors are concatenated to one vector of size $20 \times 7 \times 31 = 4,340$

Same for LBP $20 \times 7 \times 58 = 8,120$

Both vectors are normalized and concatenated to a vector $v \in R^{12460}$



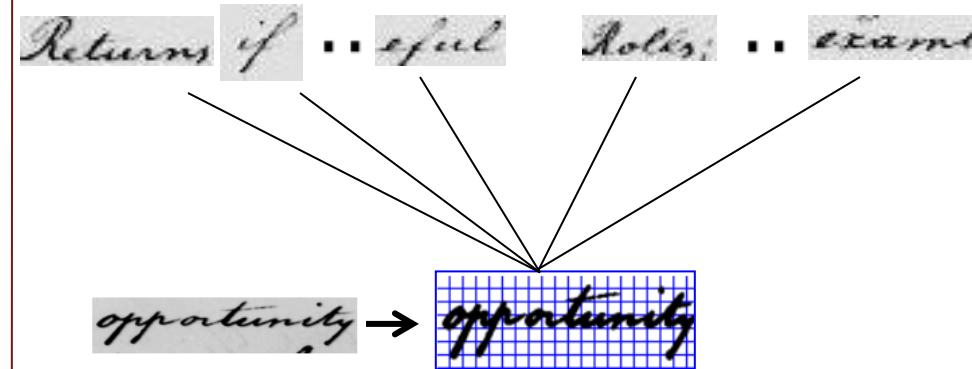
[HOG] N. Dalal and B. Triggs. **Histograms of Oriented Gradients for Human Detection**. CVPR, 2005
[LBP] T. Ojala, M. Pietikäinen, and D. Harwood. **Performance evaluation of texture measures with classification based on Kullback discrimination of distributions**. ICPR 1994.

III Similarity to prototypes

3,750 random patches are sampled and split into 250 groups of 15 patches $C \downarrow i, i=1..250$

$M \in \mathbb{R}^{12460 \times 3750}$ holds their hog+lbp vectors

A patch is first represented by its cosine similarity to the 3,750 prototypes $u = Mv$



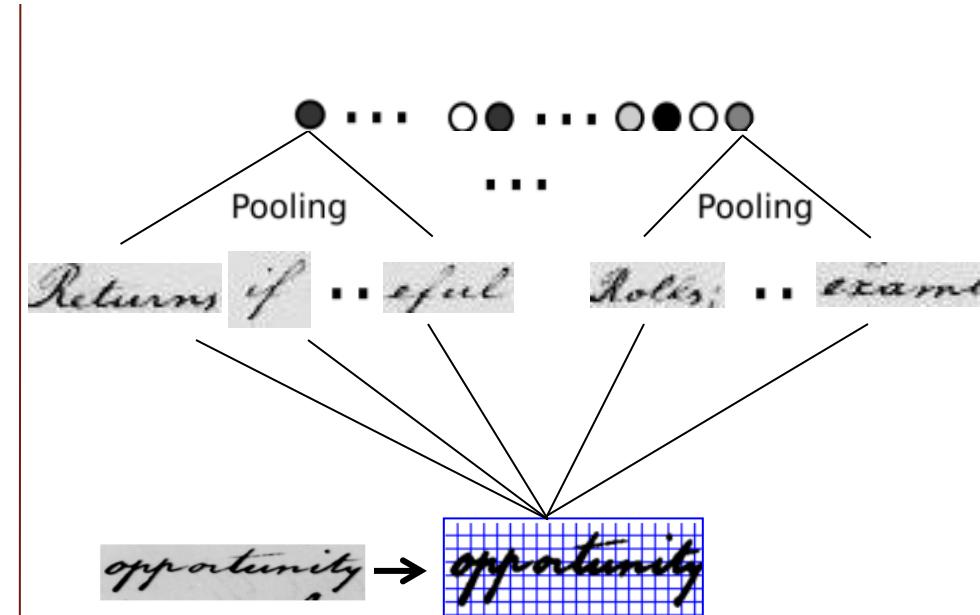
III Max pooling

3,750 random patches are sampled and split into 250 groups of 15 patches $C_i, i = 1..250$

Next, max pooling is performed over the groups C_i , resulting in $w \in R^{250}$

$$w_i = \max_{j \in C_i} u_j$$

Unlike other deep network architectures pooling here is random!

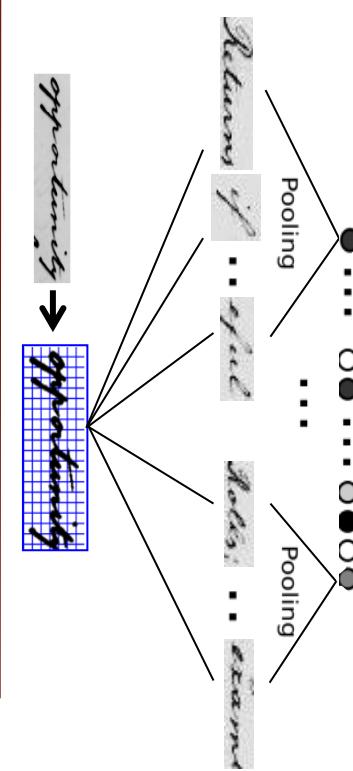


||| Query

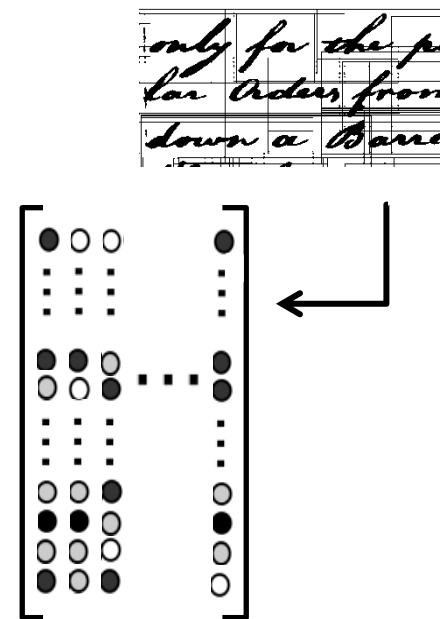
The same process is applied to the query patch and to all image patches

L₂ similarity is used to retrieve the most similar patches

Query



Dataset



Overlap Suppression

Out of all patches that share a large connected component, only the one most similar to the query is retrieved

Query: *opportunity*

*if any opportunity o
ne to be particular*

Results:

Q *opportunity*
opportunity ↗
opportunity
opportunity
Virginia. Re
xpeditionsly

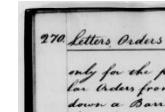
Results – examples

Q the the the the
the the the the
the the the ch.
the the the the
the this the tho.
the who the the
the the the the

Q and xan. sun
and an. sun
und as s. ons.
and me ame
and anomus
and nus. ann.
aw emma rona
ame and ans
ams rece anc
and and crew

Q October
October
October
October
October
October
October
October
little ←
at loiter ←
October.
ith bo ←

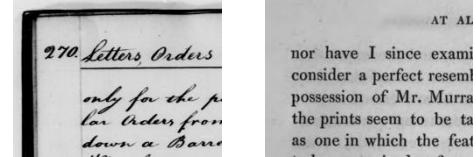
III Results – accuracy



AT AL
nor have I since examined consider a perfect resemblance of Mr. Murray's prints seem to be taken as one in which the features

	Method	GW	LB
literature	"Browsing heterogeneous document collections by a segmentation-free word spotting method" M. Rusiñol, D. Aldavert, R. Toledo, J. Lladós, ICDAR, 2011	30.5%	42.8%
	"Efficient exemplar word spotting" J. Almazán, A. Gordo, A. Fornés, E. Valveny, BMVC, 2012	54.5%	85.5%
our system	Complete pipeline	50.1%	90.7%
	Same applied to segmented words	66.3%	92.9%
	Without max-pooling ()	48.8%	90.8%
	Without max-pooling, random subset ()	47.6%	90.7%

III Results – run time



	Method	GW	LB
	Number of queries	4,860	4,988
"Efficient exemplar word spotting", Almazán et al. BMVC'12	All queries	5,058sec	4,159sec
	Average per query	1.04sec	0.83sec
Our system	All queries	158sec	46sec
	Average per query	0.03sec	0.01sec
	Single query	0.08sec	0.03sec
	Preprocessing per page	46sec	3sec
	Average memory per page	1,875KB	136KB

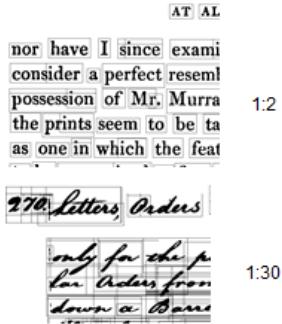
Summary

Candidate patch extraction

An elaborate heuristic based on the connected components extracted from the binary image.

Connected components are grouped to form clusters that are:

- "not too big"
- continuous in the x-axis
- components are aligned in y-axis
- contains all components in it's bounding box



HOG + LBP encoding

Each cell is represented by:

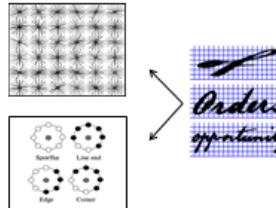
A HOG vector in R^{31}

An LBP vector in R^{58}

All HOG vectors are concatenated to one vector of size $20 \times 7 \times 31 = 4,340$

Same for LBP $20 \times 7 \times 58 = 8,140$

Both vectors are normalized and concatenated to a vector $v \in R^{12460}$



[HOG] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. CVPR, 2005
[LBP] T. Ojala, M. Pietikäinen, and D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. ICPR 1994.

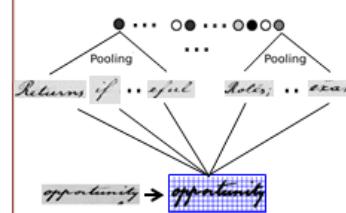
Network based representation

3,750 random patches are sampled and split into 250 groups of 15 patches $C_i, i = 1..250$

Then, the patch is represented by a vector $w \in R^{250}$ by pooling

$$w_i = \max_{j \in C_i} u_j$$

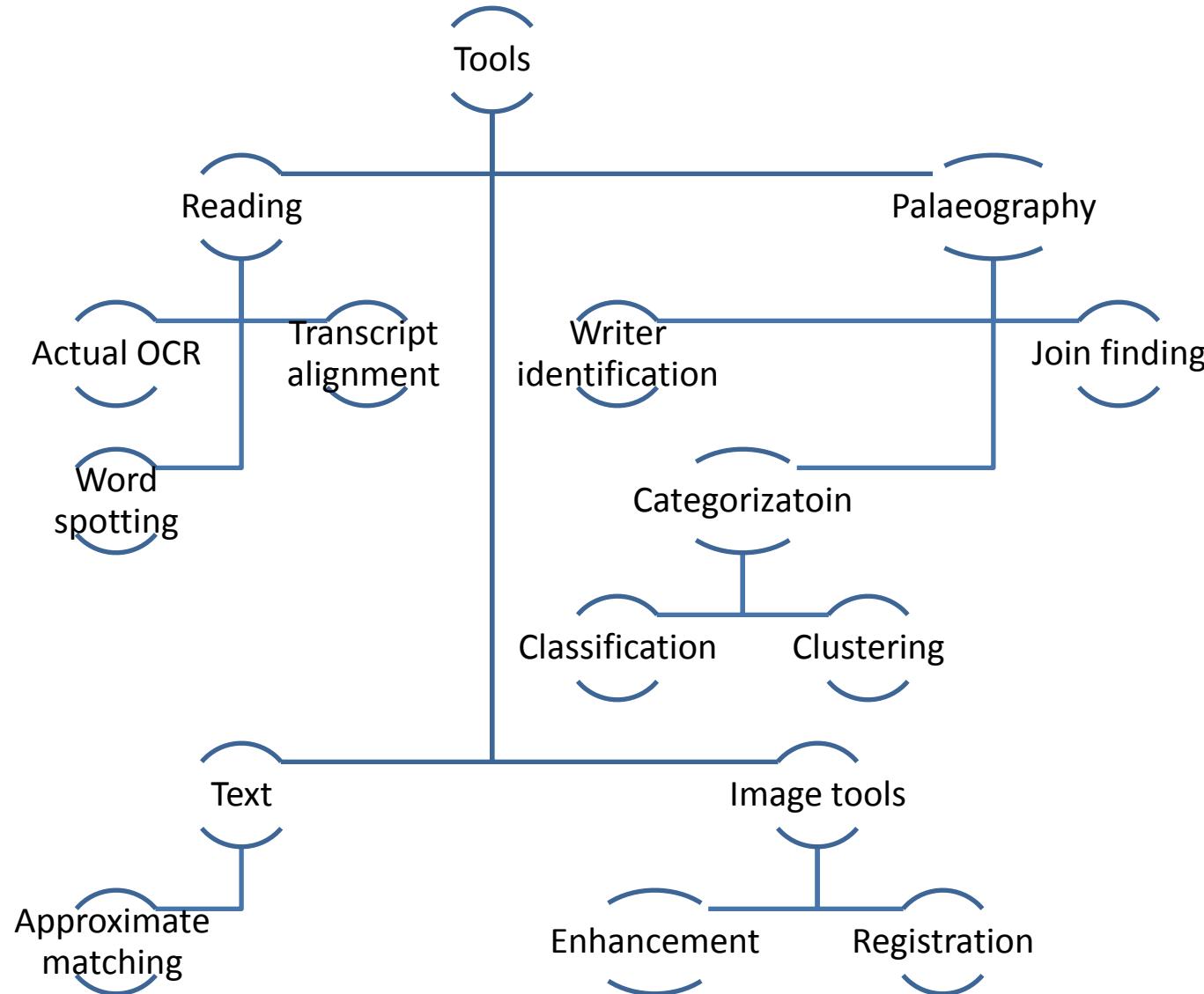
Unlike other deep network architectures pooling is random!

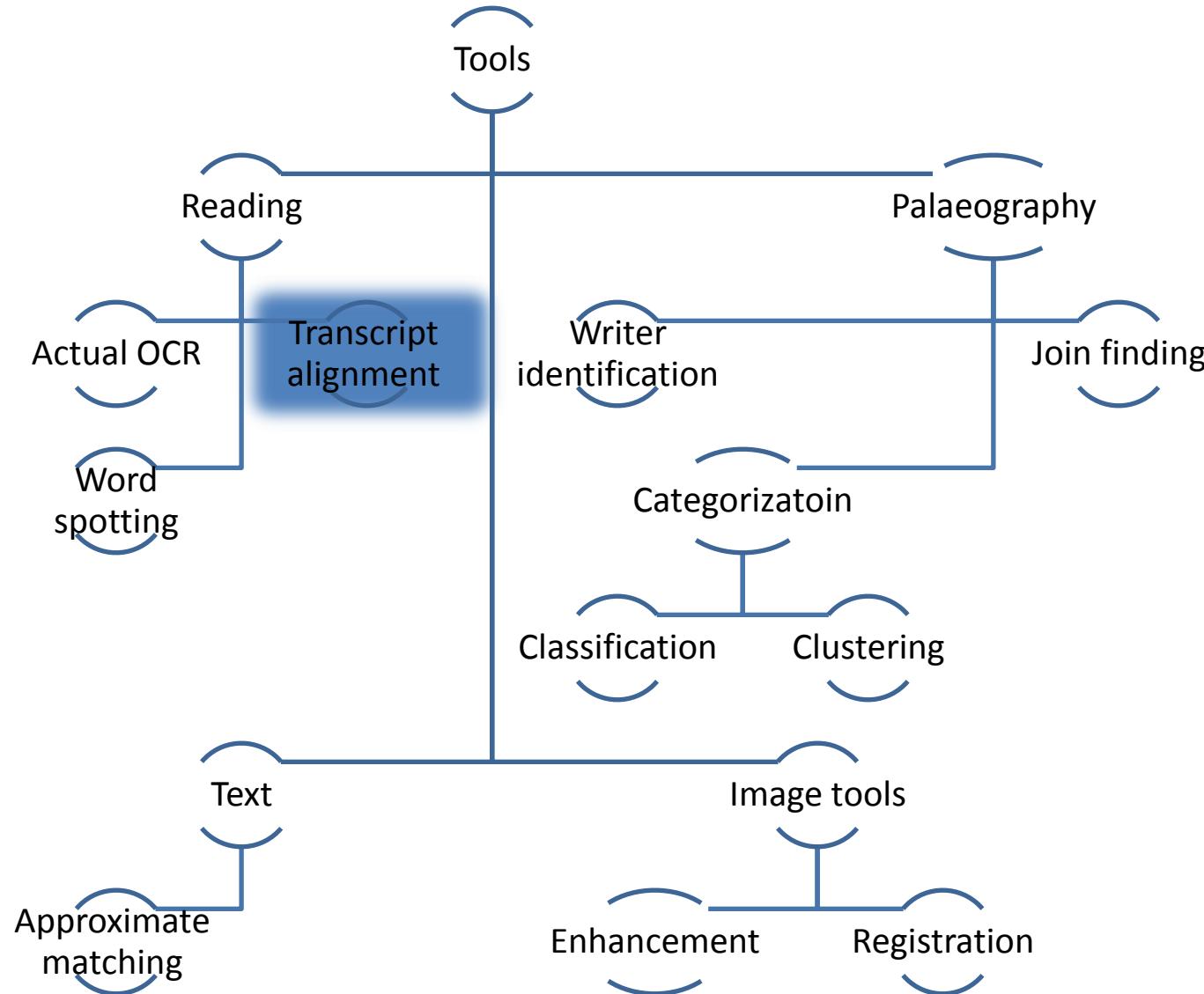


Results

	Method	Literature	
		GW	LB
M. R. Almazán et al., "Efficient exemplar word spotting", BMVC'12	"Browsing heterogeneous document collections by a segmentation-free word spotting method"	30.5%	42.8%
J. Almazán et al., "Efficient exemplar word spotting", BMVC'12	Number of queries	4,860	4,988
Sam	All queries	5,058sec	4,159sec
With	Average per query	1.04sec	0.83sec
our system	All queries	158sec	46sec
With	Average per query	0.03sec	0.01sec
With	Single query	0.08sec	0.03sec
	Preprocessing per page	46sec	3sec
	Average memory per page	1,875KB	136KB

thank
{you}





OCR Free Transcript Alignment

Image:

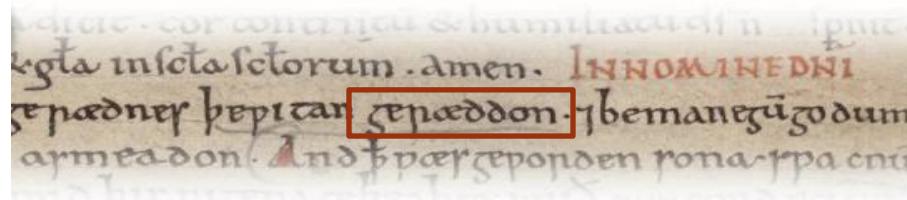
... dicit . cor contritum & humiliatum n... ipuit.
egla insciaslorum . amen . **IN NOMINE DOMINI**
þe witan geræddon . 7 be manegum g
armeadon . And þær geponden sona . 7 pa cnū

Transcription:

þe witan geræddon . 7 be manegum g
And þær wæs geworden sona swa cnút

OCR Free Transcript Alignment

Image:

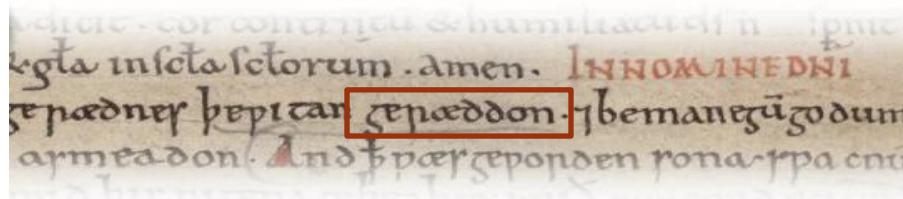


Transcription:

þe witan **geræddon**. 7 be manegum g
And þæt wæs geworden sona swa cnút

OCR Free Transcript Alignment

Image:



Transcription:

þe witan **geræddon**. 7 be manegum g
And þæt wæs geworden sona swa cnút

Motivation for alignment: the human user (UI); the computer user (word spotting)

(a)

...cūt - cor concretu & bramillacu si n...pnic
ægla insca scotorum .amen. **In nomine domini**
riproce pædnes heptan geƿæddon .ibemanegūgodum
bīnū . armeadon . And þærgeponden rona .spa enū

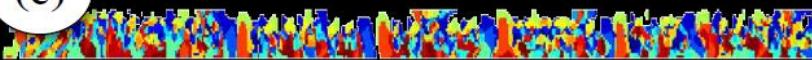
is is seo gerædnes þe witan gerædnon . 7 be manegum g
bisnum . asmeadon . And þæt wæs geworden sona swa cnút

(b)

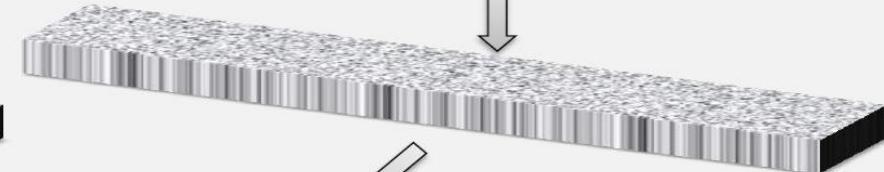
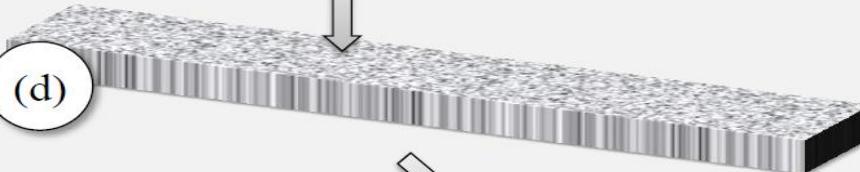
riproce pædnes heptan geƿæddon .ibemanegūgodum

is is seo gerædnes þe witan gerædnon . 7 be manegum godum

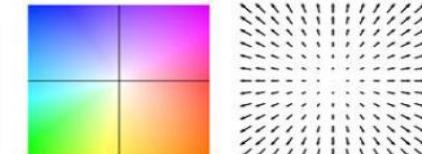
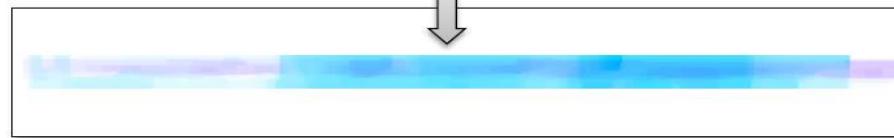
(c)



(d)



(e)



Dead Sea Scrolls

ויתרוממו למעלה לכוכב. אתה הוא יהוה בחרתך באבותינו למקדם.
וירומטו למשילה לכוכב. אתה הוא יהוה בתרתך, באבותינו למקדם.
וְהִי יְהוָה לְמַלְאָכִים לְבָנָיו וְלְבָנָתָיו בְּאֶחָתָה בְּבָרֵךְ



Genizah

אל היבנו הוא ירham עלינו וירוייה לנו מצרותינו
אל הימנו הוא ירham עלינו ידרויה לבנו מצרותינו
אהיט הוא רחט עלה וירוייה לנו מצרוועט



אסתגאת אל עזר הָדָא אֶלְי וקתרנא הָדָא לִם יַקְדֵּמָה אֶחָד אֶלְי בֵּית דִין וְלֹא שָׁהָד אֶחָד בְּחִזְרָה בֵּית
אסתגאת אל פָּצָר הָדָא אֶלְי וקתרנא הָדָא לִם יַקְדֵּמָה אֶחָד אֶל בֵּית דִין וְלֹא שָׁהָד אֶחָד בְּחִזְרָה בֵּית
פָּסְטָאָר אֲשֶׁר הָלָא כָּוָגָנוּ הָלָא כָּוָגָנוּ (על פָּסְטָאָר אֲשֶׁר חָנַצְתָּה חָטָא

Tibetan bKa gdams gsung bum

ਤੁਣਾਨਾਨਾ
ਤੁਣਾਨਾਨਾ
ਭਾਵਾਵਾਵਾ
ਭਾਵਾਵਾਵਾ
ਸਿਦਾਨਾਨਾ
ਹਾਨਾਨਾ
ਕਿਲਾਕਿਲਾ
ਕਿਲਾਕਿਲਾ
ਗਿਰਾਗਿਰਾ
ਗਿਰਾਗਿਰਾ
ਗਿਰਾਗਿਰਾ
ਗਿਰਾਗਿਰਾ

Cnut's Oxford code

is is seo gerædnes þe witan geræddon. 7 be manegum godum
is is seo gerædnes þe witan geræddon. 7 be manegum godum
is is seo gerædnes þe witan geræddon. 7 be manegum godum

Codex Sinaiticus

ΤΟΥ ΚΑΙ Ο ΤΙ ΑΝ ΠΡΟΣ

ΤΟΥ ΚΑΙ Ο ΤΙ ΑΝ ΠΡΟΣ

ΤΟΥ ΚΑΙ Ο ΤΙ ΑΝ ΠΡΟΣ