

Interpretability Digital Humanities

Oliver Eberle

I. Explainable Machine Learning - Interpretability

II. Types of interpretability

III. LRP

Theory

Practice

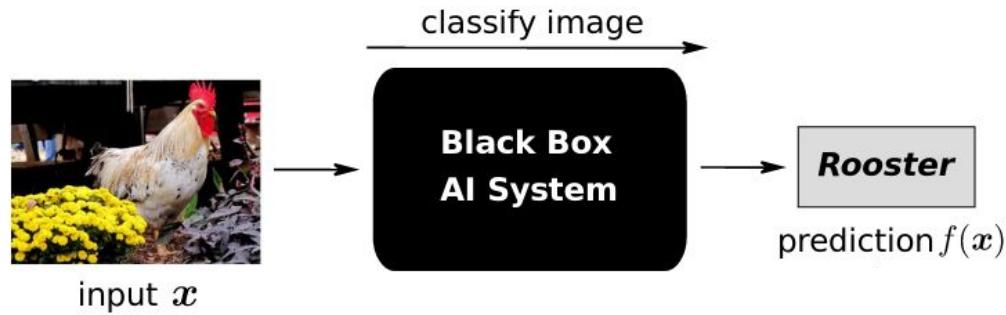
IV. Research Examples

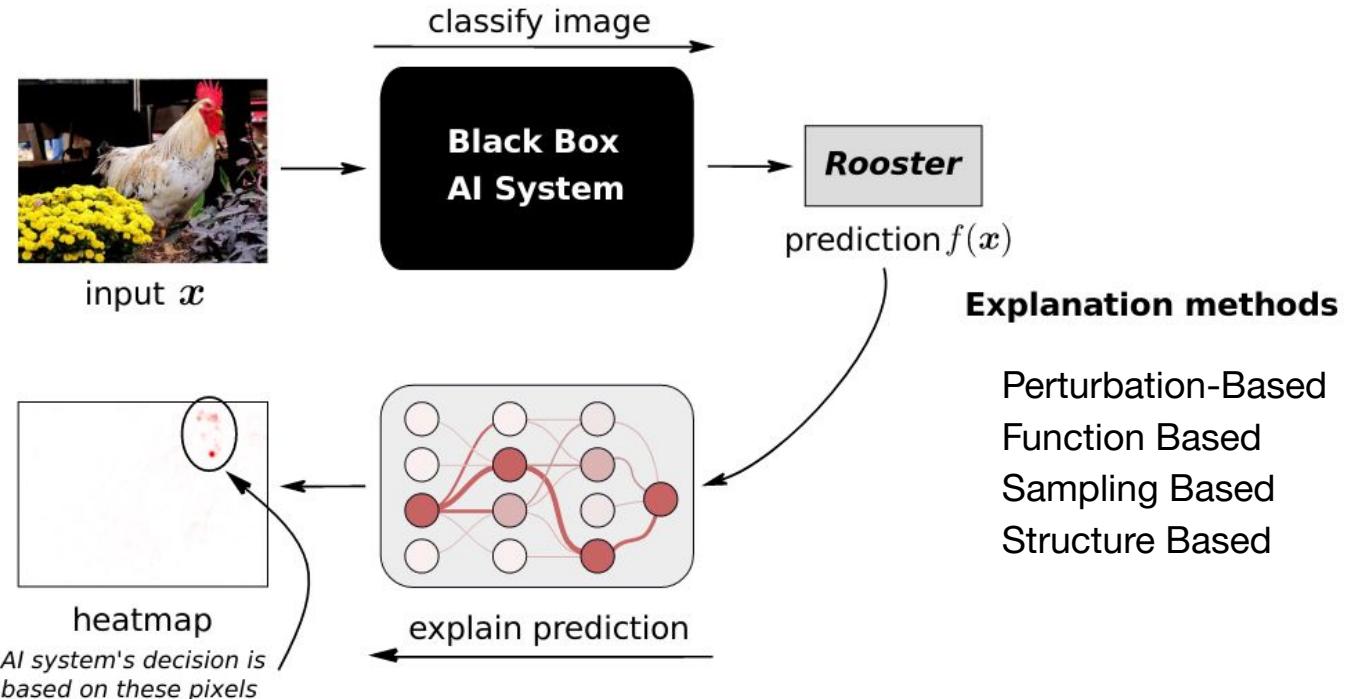
Ex1) Interpretable table similarity model

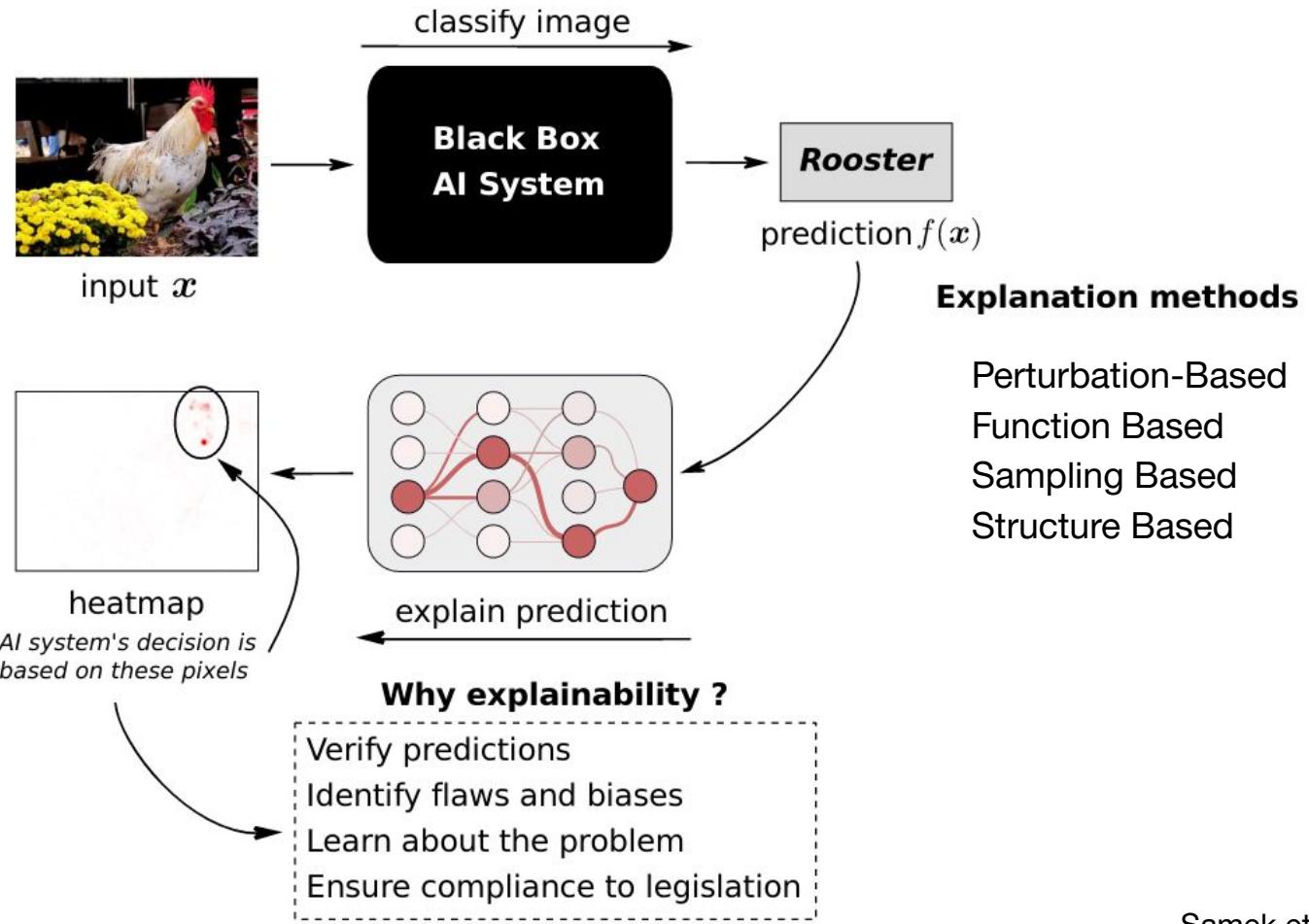
Ex2) Interpretable Graph models

IV. Discussion & Questions

I. Explainable Machine Learning - Interpretability







II. Types of interpretability

Understandability (or equivalently, intelligibility) any human can understand its function – how the model works – without expert knowledge.



Interpretability/Explainability: it is defined as the ability to explain or to provide the meaning in understandable terms to a human.



Transparency: transparent if by itself it understandable and possible to judge e.g. fairness aspects (often used in context of ethics, wide definition)



Understandability (or equivalently, intelligibility) any human can understand its function – how the model works – without expert knowledge.



Interpretability/Explainability: it is defined as the ability to explain or to provide the meaning in understandable terms to a human.

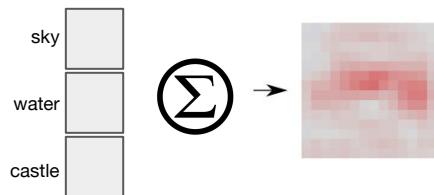
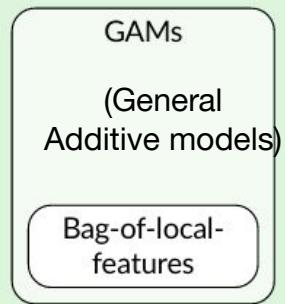
Transparency: transparent if by itself it understandable and possible to judge e.g. fairness aspects (often used in context of ethics, wide definition)

human understandability \longleftrightarrow model understandability

adapted from
Arrieta et al. (2019)

Approaches to model understandability

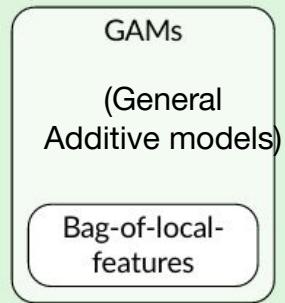
self-explainable
models



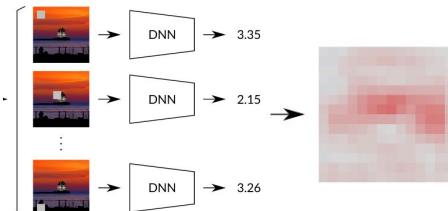
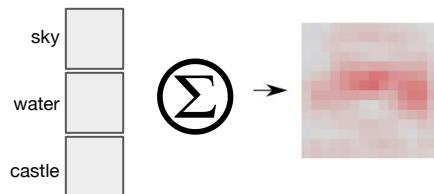
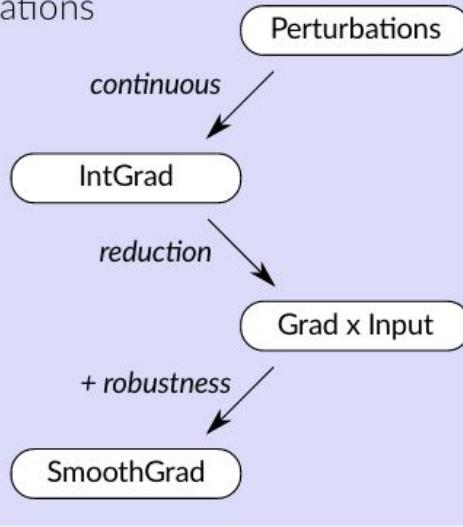
adapted from Samek & Montavon
(ECML/PKDD 2020)

Approaches to model understandability

self-explainable
models



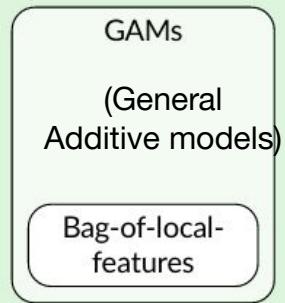
perturbation-based
explanations



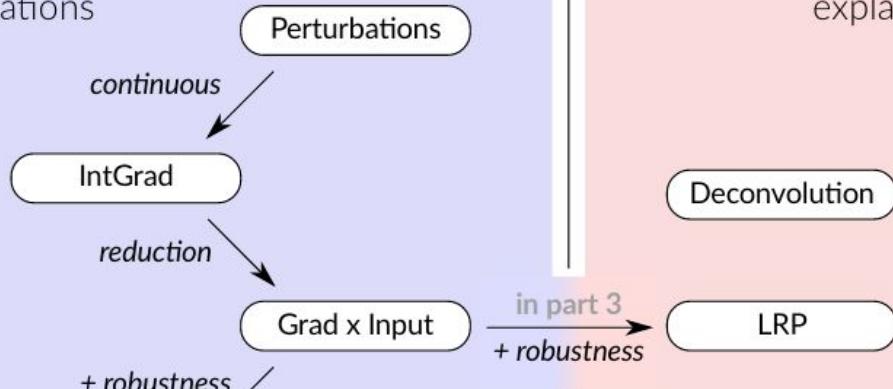
adapted from Samek & Montavon
(ECML/PKDD 2020)

Approaches to model understandability

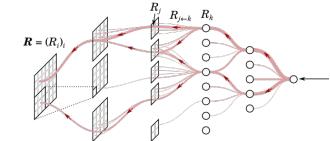
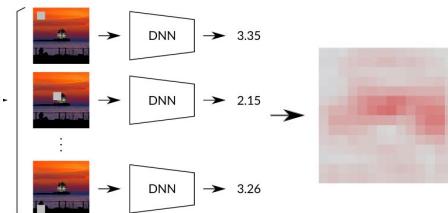
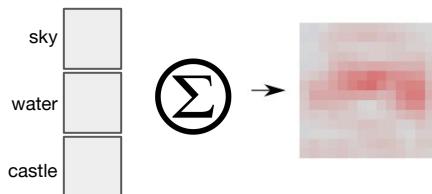
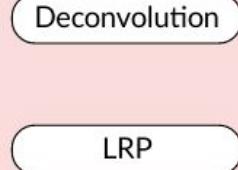
self-explainable
models



perturbation-based
explanations

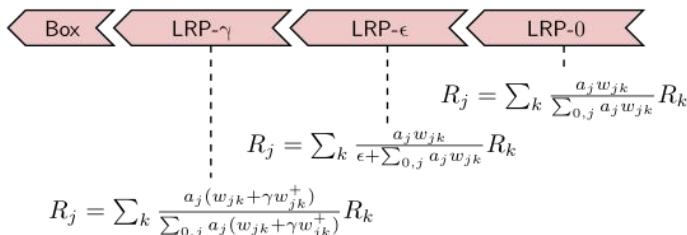
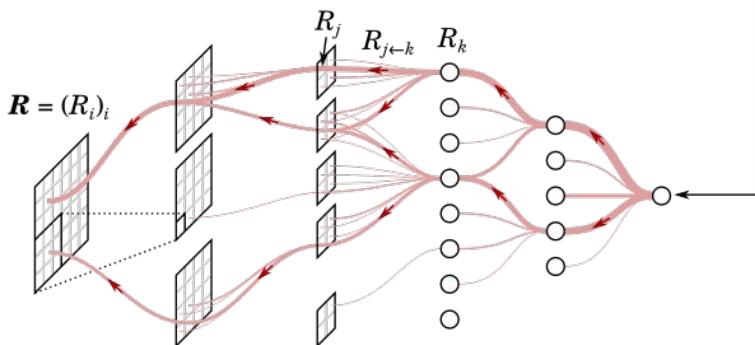
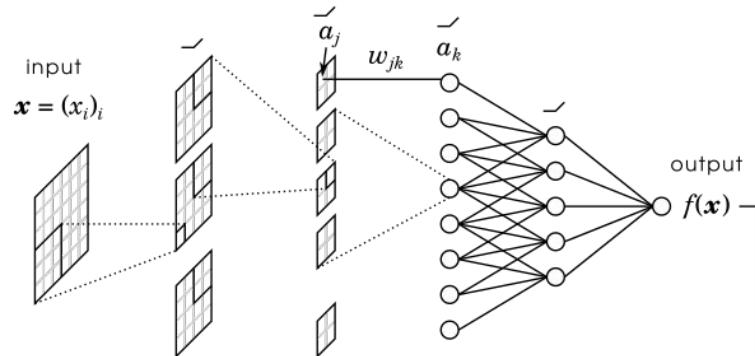


propagation-based
explanations



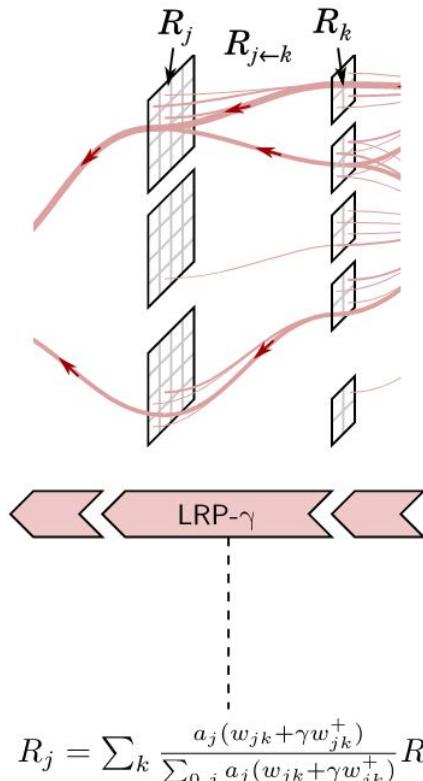
adapted from Samek & Montavon
(ECML/PKDD 2020)

III. Local Relevance Propagation



Some notation:

- ▶ j and k : neurons from successive layers
- ▶ w_{jk} : weight connecting neuron j to neuron k
- ▶ w_{0k} : bias for neuron k .
- ▶ $\sum_{0,j}$ sum over all input neurons j of neuron k and the bias.
- ▶ ReLU neuron: $a_k = \max(0, \sum_{0,j} a_j w_{jk})$.



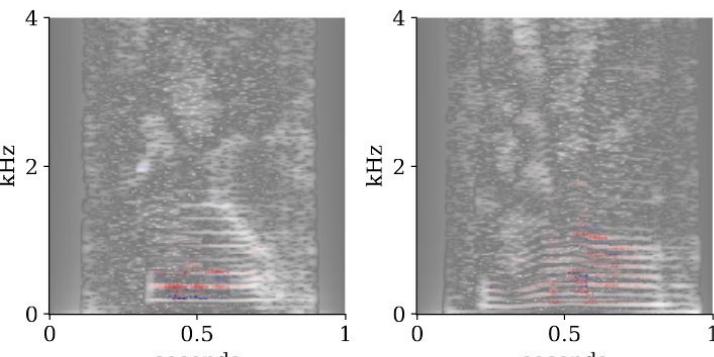
Example: LRP-γ

$$R_j = \sum_k \frac{a_j(w_{jk} + \gamma w_{jk}^+)}{\sum_{0,j} a_j(w_{jk} + \gamma w_{jk}^+)} R_k$$

- ▶ $a_j(w_{jk} + \gamma w_{jk}^+)$: Contribution of neuron a_j to the activation a_k .
- ▶ R_k ‘Relevance’ of neuron k available for redistribution.
- ▶ $\sum_{0,j} a_j(w_{jk} + \gamma w_{jk}^+)$ Normalization term that implements conservation.
- ▶ \sum_k : Pool all ‘relevance’ received by neuron j from the layer above.

++	19.	a worthy entry into a very difficult genre .
	20.	it 's a good film -- not a classic , but odd , entertaining and authentic .
--	21.	it never fails to engage us .

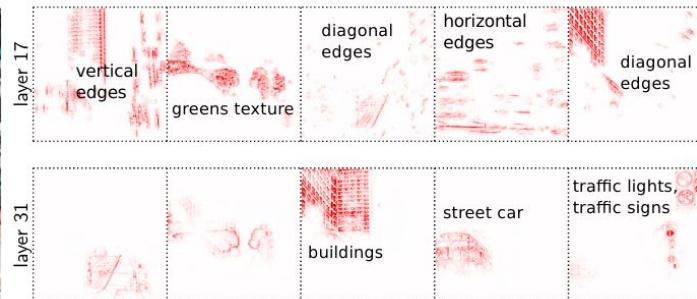
Arras et al. (2017)



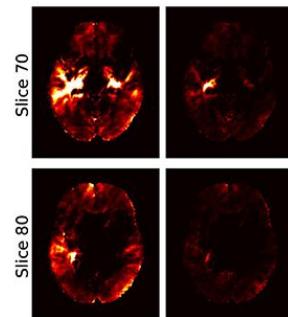
(c) female speaker, zero

(d) male speaker, zero

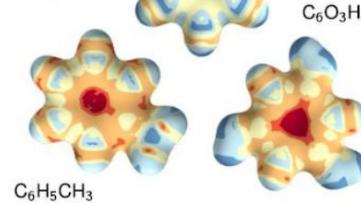
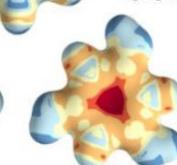
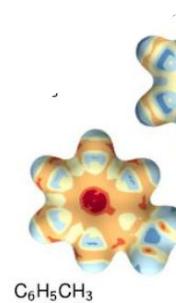
Becker et al.
(2018)



Kauffmann et al. (2019)



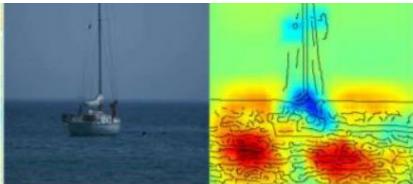
Böhle et al.
(2019)



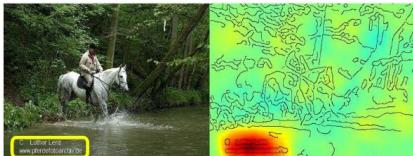
Schütt et al. (2017,
2018)



The tale of Clever Hans....



Horse-picture from Pascal VOC data set



Source tag
present
↓

Classified
as horse

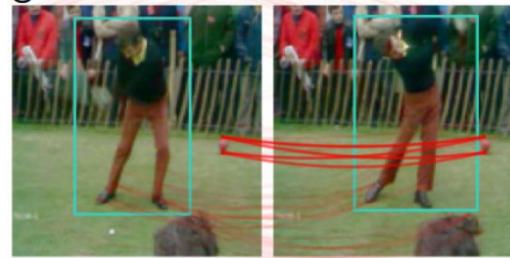
No source
tag present
↓

Not classified
as horse

soccer



golf



Dragonfly



Banana

Hendrycks et al.
(2019)

Eberle et al.
(2020)

Lapuschkin et al.
(2019)

Woody Allen 's best picture.

IV. Research Examples

Understanding Representations using Explanations of ...

Ex1) Second-Order in Similarity (BiLRP)



Ex2) Higher-Order in GNNs (GNN-LRP)



Example 1) Interpretable table similarity model

Gradi	minu			T A B.			pt.
	d ^o .	o	i	pt.	m	z	
0	0	0	0	1	2	50	2
1	0	1	3		3	53	
2		2	6		4	55	
3		3	9		5	58	
4		4	11		7	1	
5		5	14		8	4	
6		6	17		9	7	
7		7	20		10	9	

Gradi	minu			T A V			pt.
	o	i	v	pt.	m	z	
0	0	0	0	1	2	50	2
1	0	1	3		3	53	
2		2	6		4	55	
3		3	9		5	58	
4		4	11		7	1	
5		5	14		8	4	
6		6	17		9	7	
7		7	20		10	9	

Example 1) Interpretable table similarity model

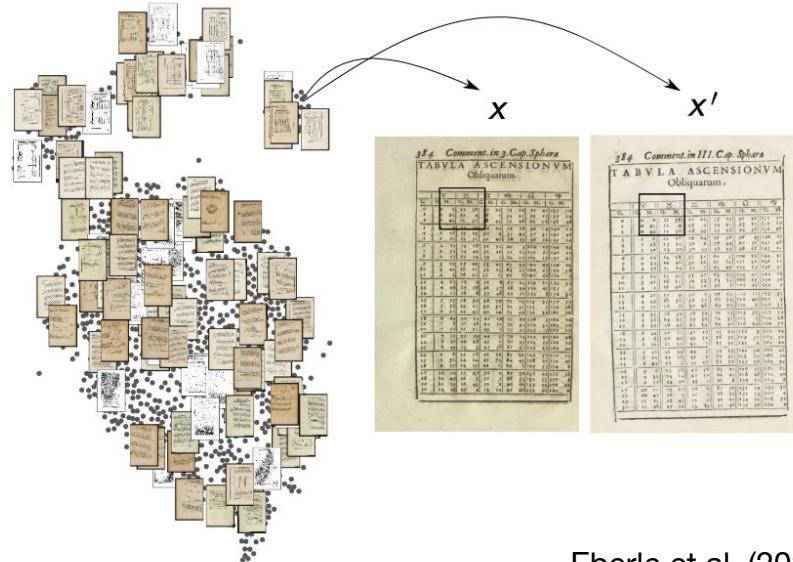
Challenges

Data

- Inhomogeneous data (text/ numbers/ geometry/ diagrams,...)
- Costly annotation process

Modelling

- OCR approaches fail to extract historical text
- Geometric row-column structure
- Infinite vocabulary size
- Pixel-vs-semantic similarity
- Explainable table representations



Eberle et al. (2020)

Table Embeddings

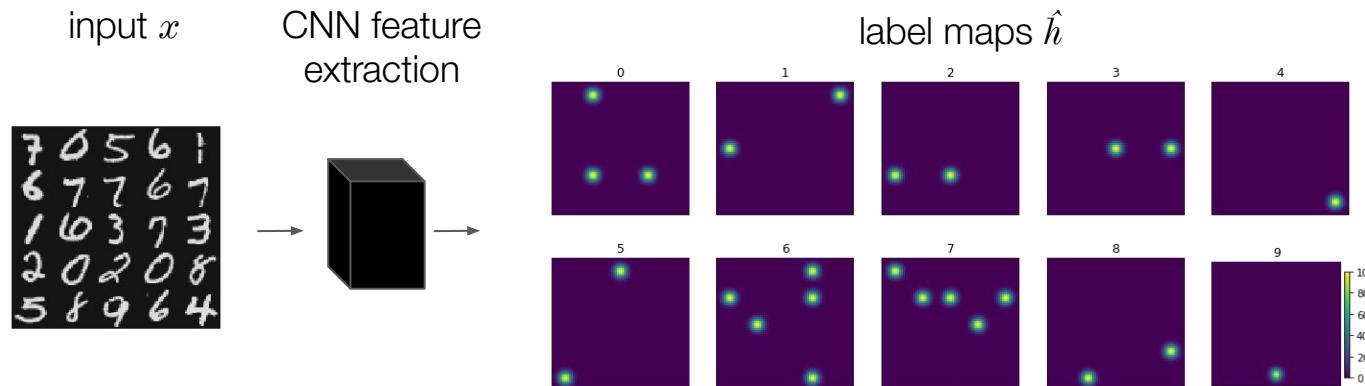


Table Embeddings - annotated data

P A R S
 Tabula declinationis Solis, cuiuslibet
 gradus Eclipticae maxima declinatione
 23. graduum, 28. minutorum
 suffutata.

Nod.	Libra.	Scorpi.	Sagit.
odi.	Aries.	Taurus.	Geniini
G.	G. M. 2.	G. M. 2.	G. M. 2. G.
0	0 0 0	11 29 5	20 20 15 30
1	0 23 5;	11 50 6	20 22 17 29
2	0 47 46	12 10 56	10 5 7 28
3	1 11 39	12 31 34	20 15 55 27
4	1 15 30	12 51 59	0 58 20 26
5	1 59 20	13 12 22 21	9 21 25
6	2 23 8;	13 32 12 21	19 59 24
7	2 40 54;	13 51 58	21 30 19 23
8	2 10 37 14	21 50 21 40	5 22
9	3 34 18 14	30 48 21 46	19 28
10	3 57 5 14	49 51 21 53	29 20
11	3 21 28 15	8 40 22 7	6 19
12	4 44 5 15	27 13 22 15	17 18
Virgo.	Leo.	Cancer.	
Pisces.	Aquar.	Capric.	

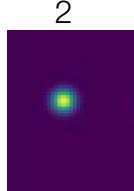
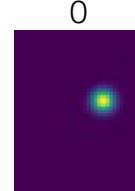
Per 30. Quadrantis Minuta.

Arcus. G. M.	Sinus.	Arcus. G. M.	Sinus.
42 30	6755902	54 30	8141155
43 0	6829984	55 0	8191520
43 30	6883546	55 30	8241262
44 0	6946584	56 0	8290376
44 30	7009093	56 30	8338858
45 0	7871068	57 0	8386706
45 30	5132504	57 30	8433915
46 0	7193398	58 0	8480481
46 30	7253744	58 30	8526402
47 0	7313537	59 0	8571673
47 30	13727773	59 30	8616292
48 0	7431448	60 0	8660257
48 30	7489557	60 30	8703557
49 0	7547096	61 0	8746197
49 30	7604060	61 30	8788111
50 0	7660445	62 0	8829476
50 30	7716246	62 30	8870108
51 0	7771460	63 0	8910065
51 30	7826082	63 30	8949344
52 0	7880108	64 0	8987940
52 30	7933533	64 30	9025853
53 0	7986355	65 0	9062078
53 30	80381569	65 30	9099613
54 0	8090170	66 0	9135155

Representing compositions

9	6	9
1		
2	20	22
2	21	24
3	22	25

input x

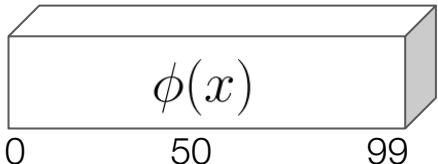


Modelling table pairs

input

x

9	6	9
1	20	22
2	21	24
3	22	25



summed
evidence

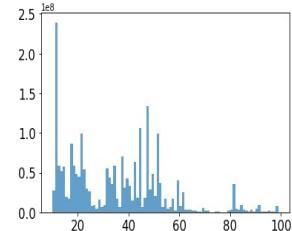
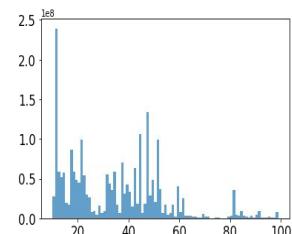
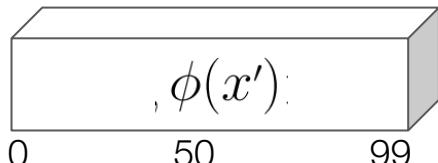


table
similarity

0.71

x'

22	21	53
24	22	54
25	23	54
26	24	55

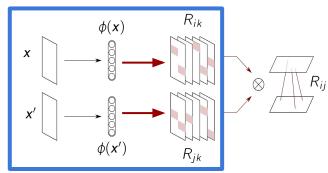


$$y = \langle \phi(x), \phi(x') \rangle$$

$$\phi_m(x) = \min(h_i, h_j * \delta)$$

$$s_m = \sum_i \phi_m(x_i)$$

I. Explaining table similarity



feature map explanations ϕ_m

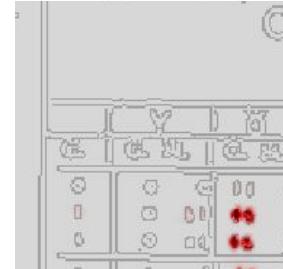
$m=11$

$m=12$

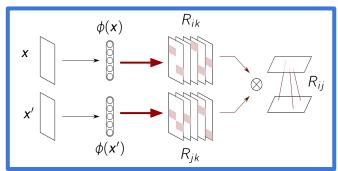
	V		ꝝ
	G.	M.	G.
	0	0	11
0	22	12	
0	44	12	
	1	6	12



	V		ꝝ
	G.	M.	G.
	0	0	11
0	22	12	
0	44	12	
	1	6	12



I. Explaining table similarity



feature map explanations ϕ_m

$m=11$

	V		λ
	G.	M.	G.
	0	0	11
	0	22	12
	0	44	12
	1	6	12

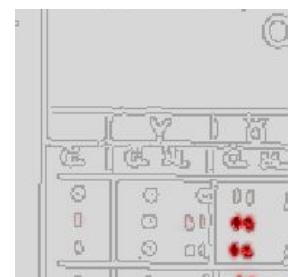


$m=12$

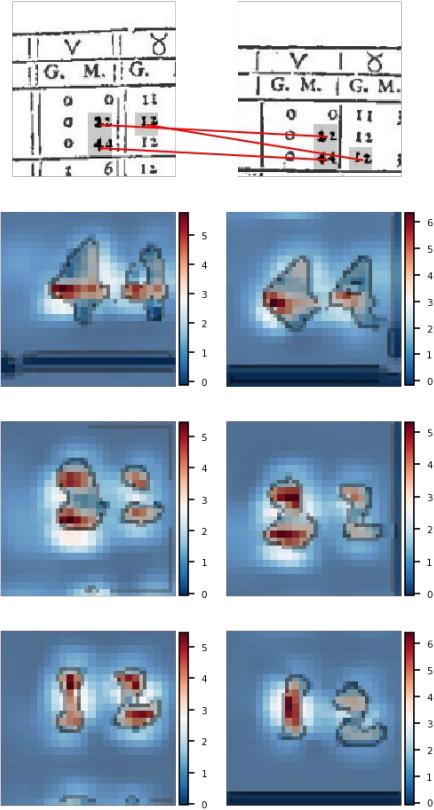


patch-level
explanations

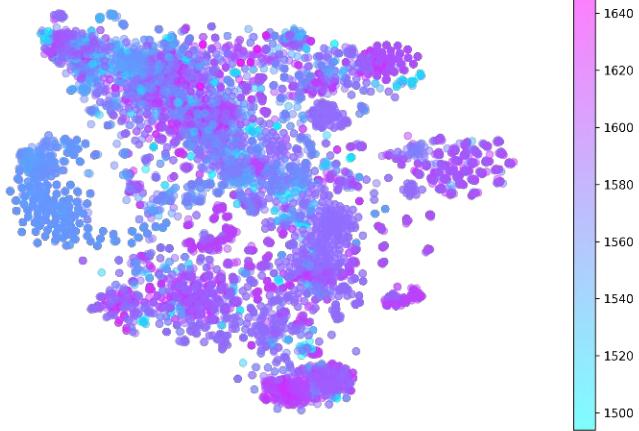
	V		λ
	G.	M.	G.
	0	0	11
	0	22	12
	0	44	12
	1	6	12



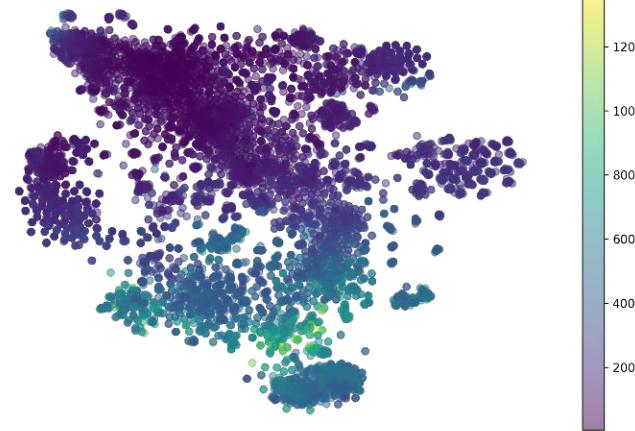
pixel-level
explanations



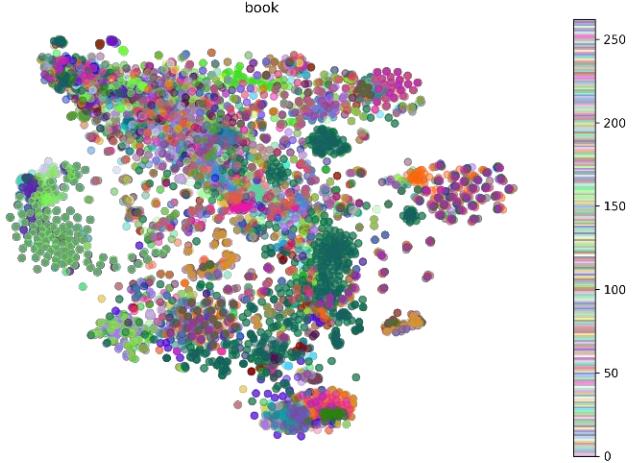
print year



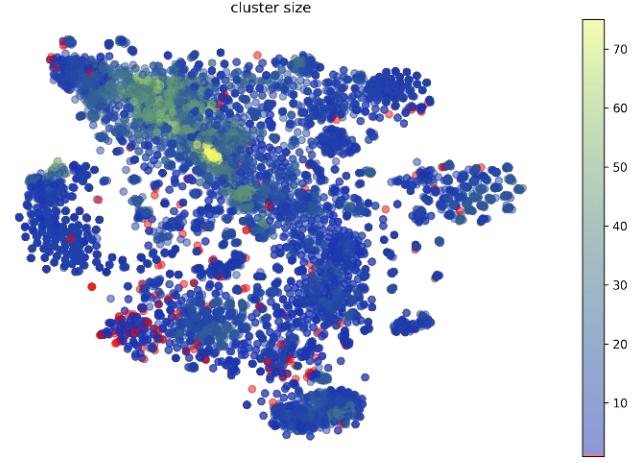
digit density



book



cluster size



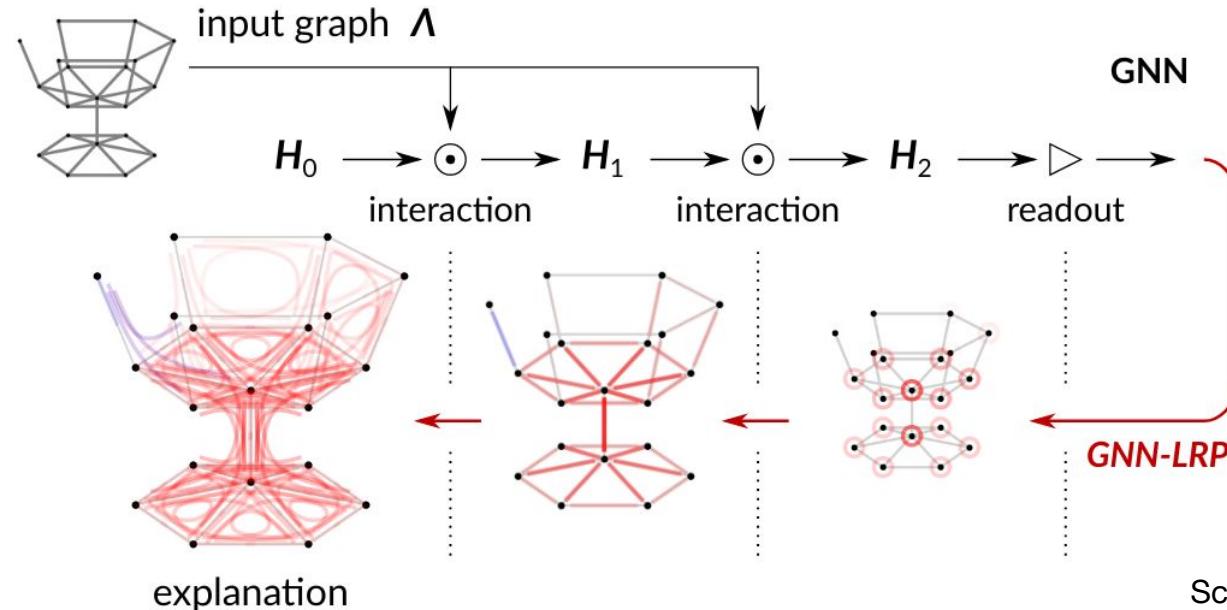
Example 2) Interpretable Graph models

GNN
mechanism

$$\text{aggregate: } \mathbf{Z}_t = \boldsymbol{\Lambda} \mathbf{H}_{t-1}$$

$$\text{combine: } \mathbf{H}_t = (\mathcal{C}_t(\mathbf{Z}_{t,K}))_K$$

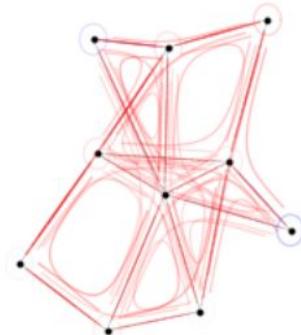
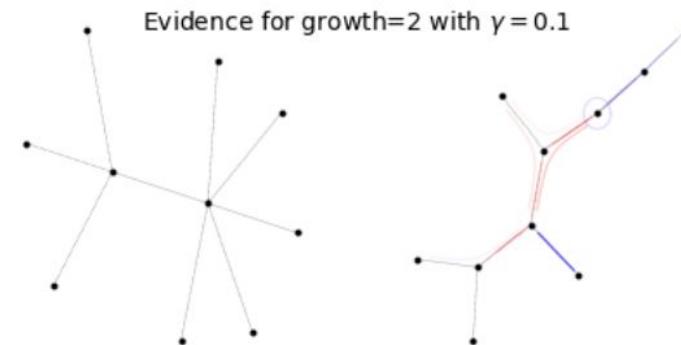
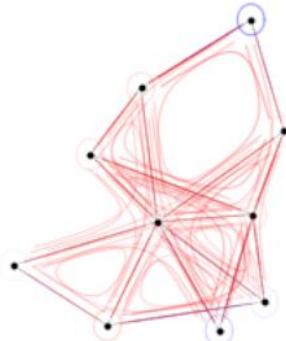
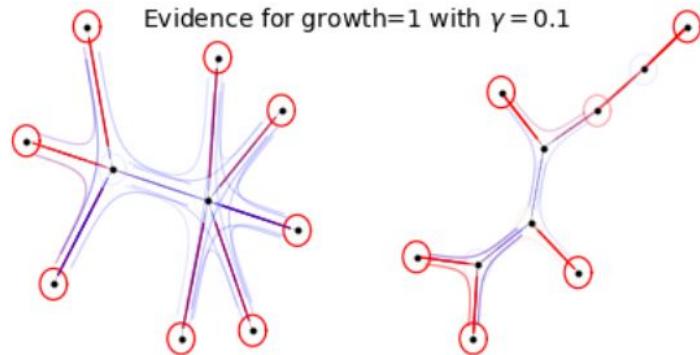
Making GNNs
interpretable



Schnake et al. (2020)

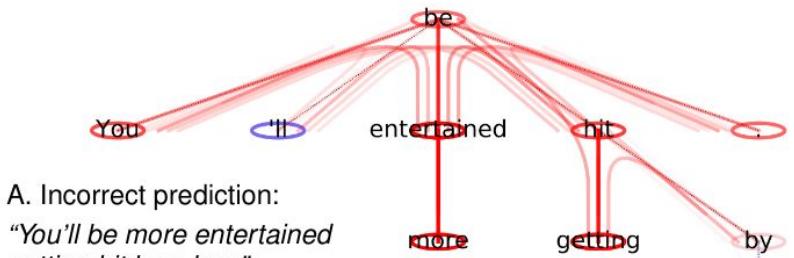
Example 2) Interpretable Graph models

Synthetic graphs



Example 2) Interpretable Graph models

Sentiment GNN



A. Incorrect prediction:

"You'll be more entertained
getting hit by a bus."

→ **positive**



B. Entity bias:

"Hugh Grant and Sandra
Bullock are two such
likeable actors." → **positive**

IV. Discussion & Questions

Challenges

- Humanities Researcher / Machine Learning Scientist
- Concepts of input data and source material
- Automation, pipelines and statistical analyses
- Opening the black box
- The chances and challenges of transfer learning

...

Table Embeddings

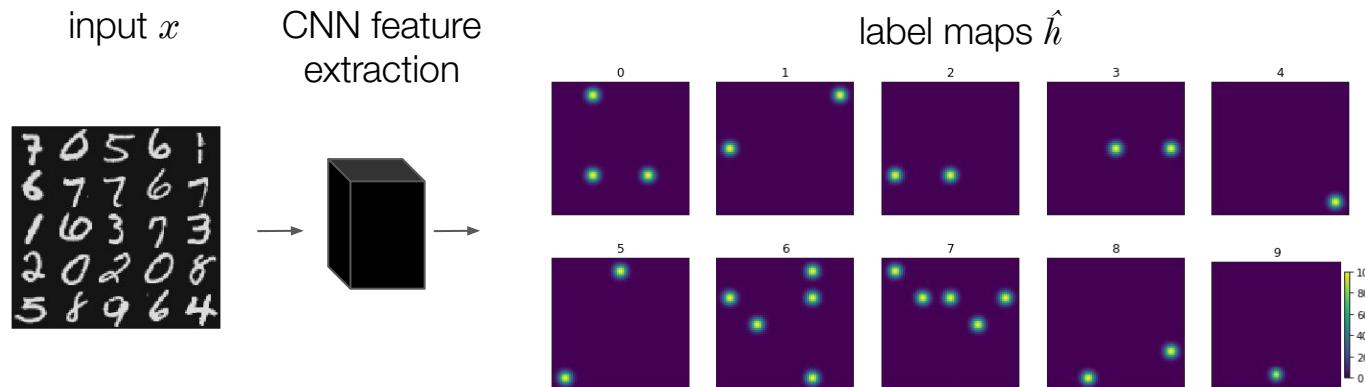


Table Embeddings - annotated data

P A R S
*Tabula declinationis Solis, cuiuslibet
gradus Eclipticae maxima declinatione
23 graduum, 28. minutorum
supputata.*

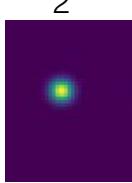
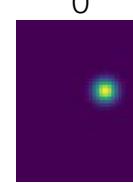
N	Libra.	Scorpi.	Sagit.	
odi.	Aries.	Taurus.	Geniini	
G.	G.M. 2.	G.M. 2.	G.M. 2. G.	
0	0 0 0	11 29 5	20 20 15	30
1	0 23 5;	11 50 6	20 22 17	29
2	0 47 46	12 10 56	10 5 7	28
3	1 11 39	12 31 34	20 15 55	27
4	1 15 36	12 51 59	0 58 20	26
5	1 59 20	13 12 22	21 9 21	25
6	2 23 8;	13 32 12	21 19 59	24
7	2 40 54;	13 51 58	21 30 19	23
8	2 10 37 14	21 30 21	40 5 22	
9	3 34 18 14	30 48 21 46	19 2 21	
10	3 57 5 14	49 51 21 53	19 20	
11	3 21 28 15	8 40 22 7	6 19	
12	4 44 5 15	27 13 22 15	17 18	
Virgo.	Leo.	Cancer.		
Pisces.	Aquar.	Capric.		

Per 30. Quadrantis Minuta.							
Arcus.		Sinus.		Arcus.		Sinus.	
G.	M.			G.	M.		
42	30	6755902		54	30	8141155	
43	0	6829984		55	0	8191520	
43	30	6883546		55	30	8241262	
44	0	6946584		56	0	8290376	
44	30	7009093		56	30	8338858	
45	0	7871068		57	0	8386706	
45	30	5132504		57	30	8433915	
46	0	7193398		58	0	8480481	
46	30	7253744		58	30	8526402	
47	0	7313537		59	0	8571673	
47	30	1372773		59	30	8616292	
48	0	7431448		60	0	8660257	
48	30	7489557		60	30	8703557	
49	0	7547096		61	0	8746197	
49	30	7604060		61	30	8788111	
50	0	7660445		62	0	8829476	
50	30	7716246		62	30	8870108	
51	0	7771460		63	0	8910065	
51	30	7826082		63	30	8949344	
52	0	7880108		64	0	8987940	
52	30	7933533		64	30	9025853	
53	0	7986355		65	0	9062078	
53	30	8038159		65	30	9099613	
54	0	8090170		66	0	9135155	

Representing compositions

0	0	0
I	20	22
2	21	24
3	22	25

input x



Detecting single numbers

i = 0

	60	61	62	63
0	8660254	8746197	8829476	8910065
1	8661708	8747607	8830841	8911385
2	8663162	8749016	8832205	8912704
3	8664615	8750425	8833569	8914023
4	8666067	8751833	8834932	8915341
5	8667518	8753240	8836295	8916659
6	8668968	8754646	8837657	8917976
7	8670487	8756051	8839018	8919292
8	8671866	8757456	8840378	8920607
9	8673314	8758860	8841737	8921921
10	8674762	8760263	8843095	8923234
11	8676209	8761665	8844452	8924546
12	8677655	8763068	8845809	8925858
13	8679100	8764468	8847105	8927169
14	8680544	8765868	8848521	8928479
15	8681988	8767267	8849876	8929789

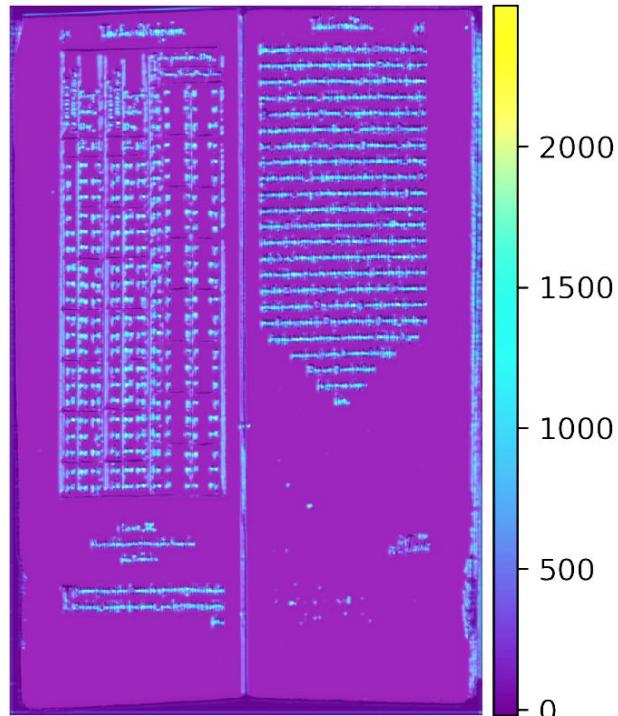
i = 7

	60	61	62	63
0	8660254	8746197	8829476	8910065
1	8661708	8747607	8830841	8911385
2	8663162	8749016	8832205	8912704
3	8664615	8750425	8833569	8914023
4	8666067	8751833	8834932	8915341
5	8667518	8753240	8836295	8916659
6	8668968	8754646	8837657	8917976
7	8670487	8756051	8839018	8919292
8	8671866	8757456	8840378	8920607
9	8673314	8758860	8841737	8921921
10	8674762	8760263	8843095	8923234
11	8676209	8761665	8844452	8924546
12	8677655	8763068	8845809	8925858
13	8679100	8764468	8847105	8927169
14	8680544	8765868	8848521	8928479
15	8681988	8767267	8849876	8929789

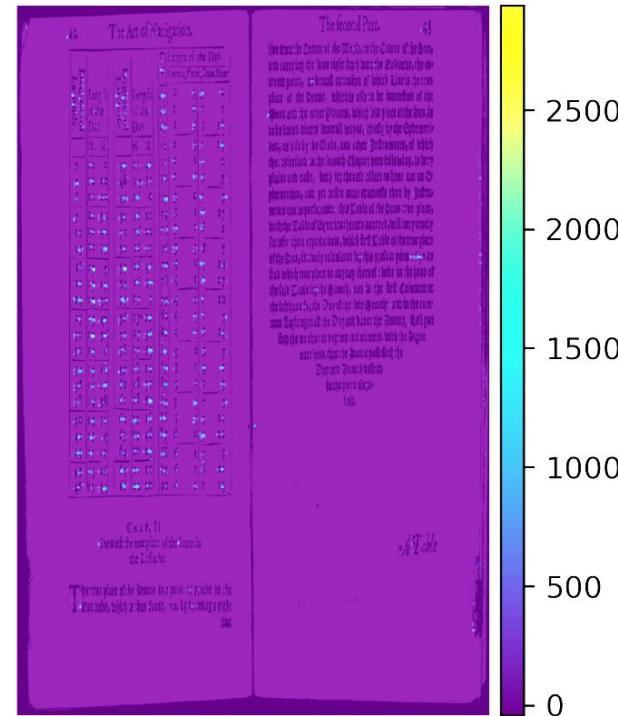
How to learn sets or compositions of numbers?

Improve detection using adversarial text samples

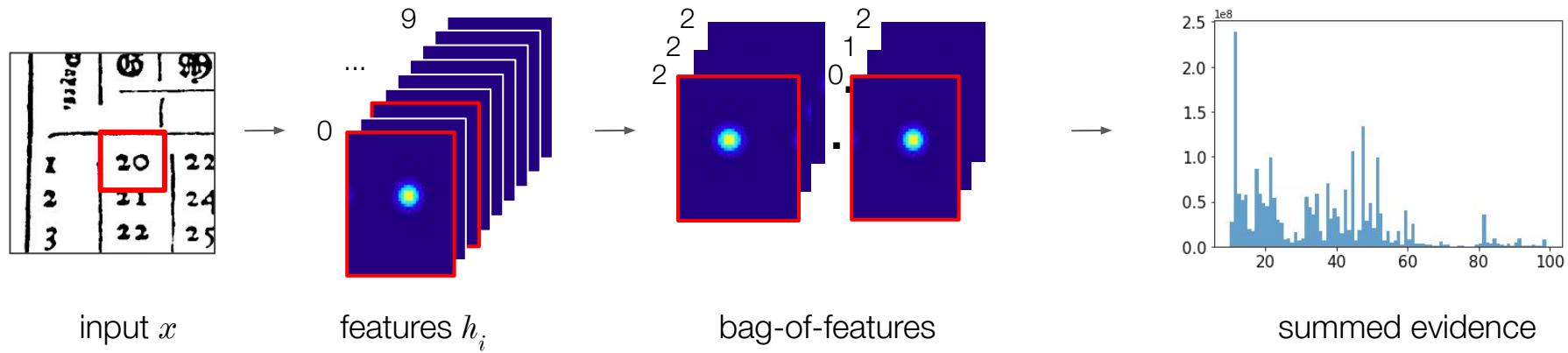
adv_False_pret_True.pt



adv_True_pret_True.pt



Modelling composition of numbers



$$\phi_m(x) = h_i(x)h_j(x * \delta)$$

or

$$\phi_m(x) = \min(h_i, h_j * \delta)$$

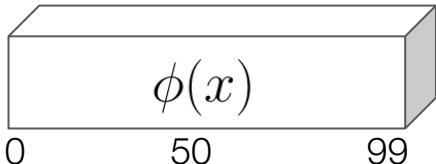
$$s_m = \sum_i \phi_m(x_i)$$

Modelling table pairs

input

x

9	6	9
1	20	22
2	21	24
3	22	25



summed
evidence

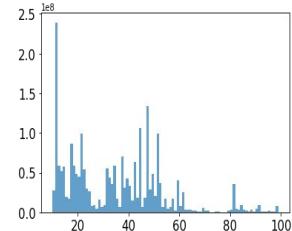
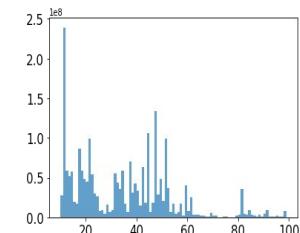
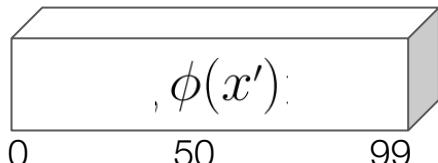


table
similarity

0.71

x'

22	21
24	22
25	23
26	24



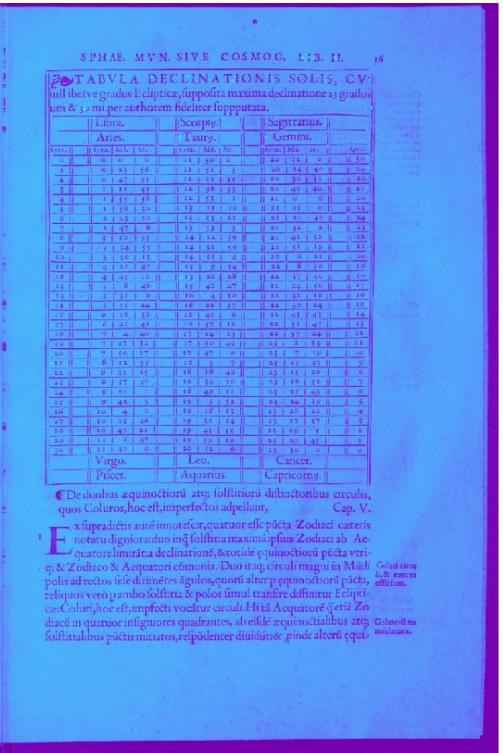
$$y = \langle \phi(x), \phi(x') \rangle$$

$$\phi_m(x) = \min(h_i, h_j * \delta)$$

$$s_m = \sum_i \phi_m(x_i)$$

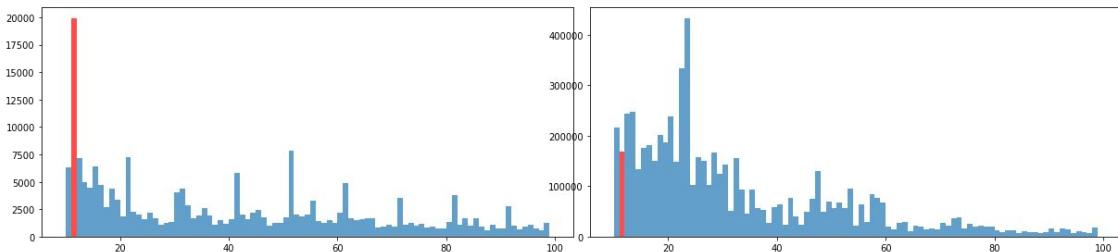
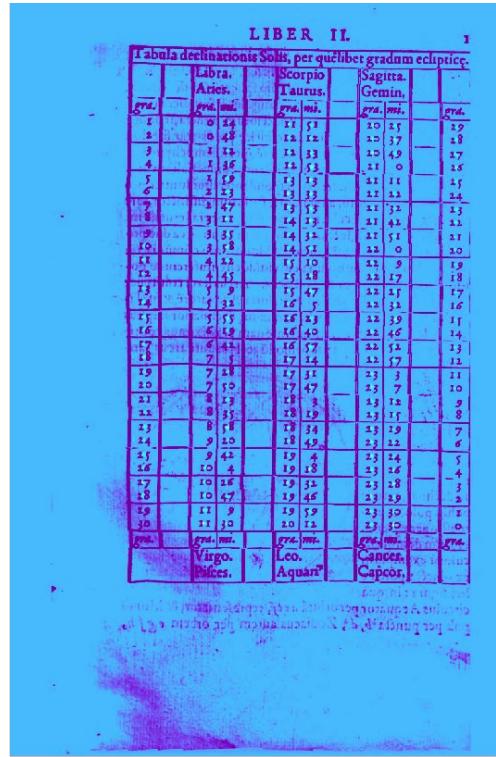
Most similar

visually different!

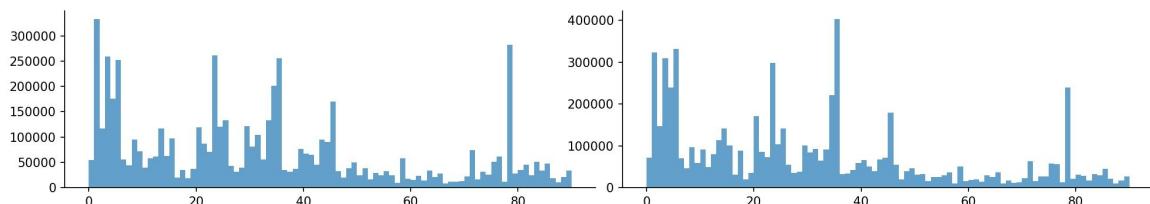
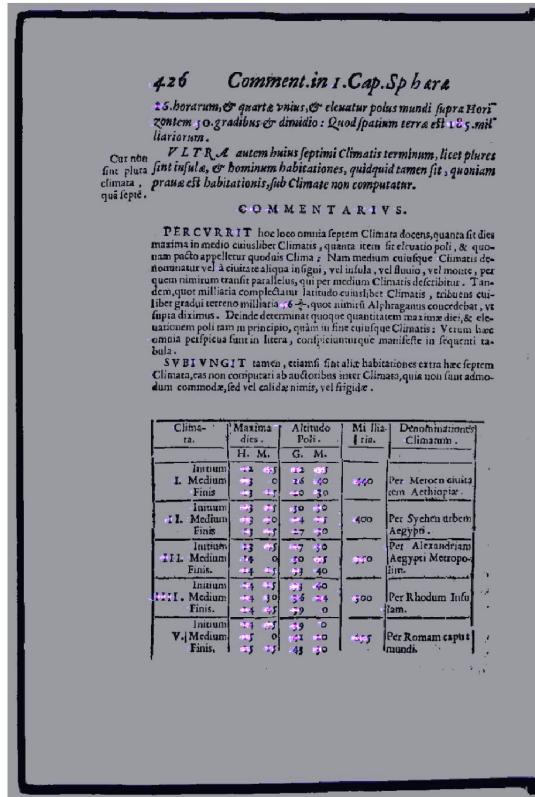
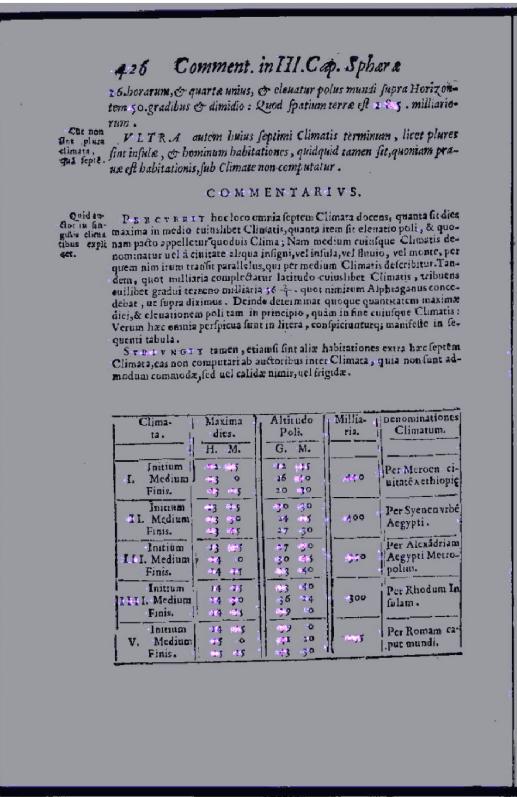


Etiam aquinoctiora aut fulliniora distinctionibus circulis, quia Colores hoc eff. imperfectos adpelluntur.

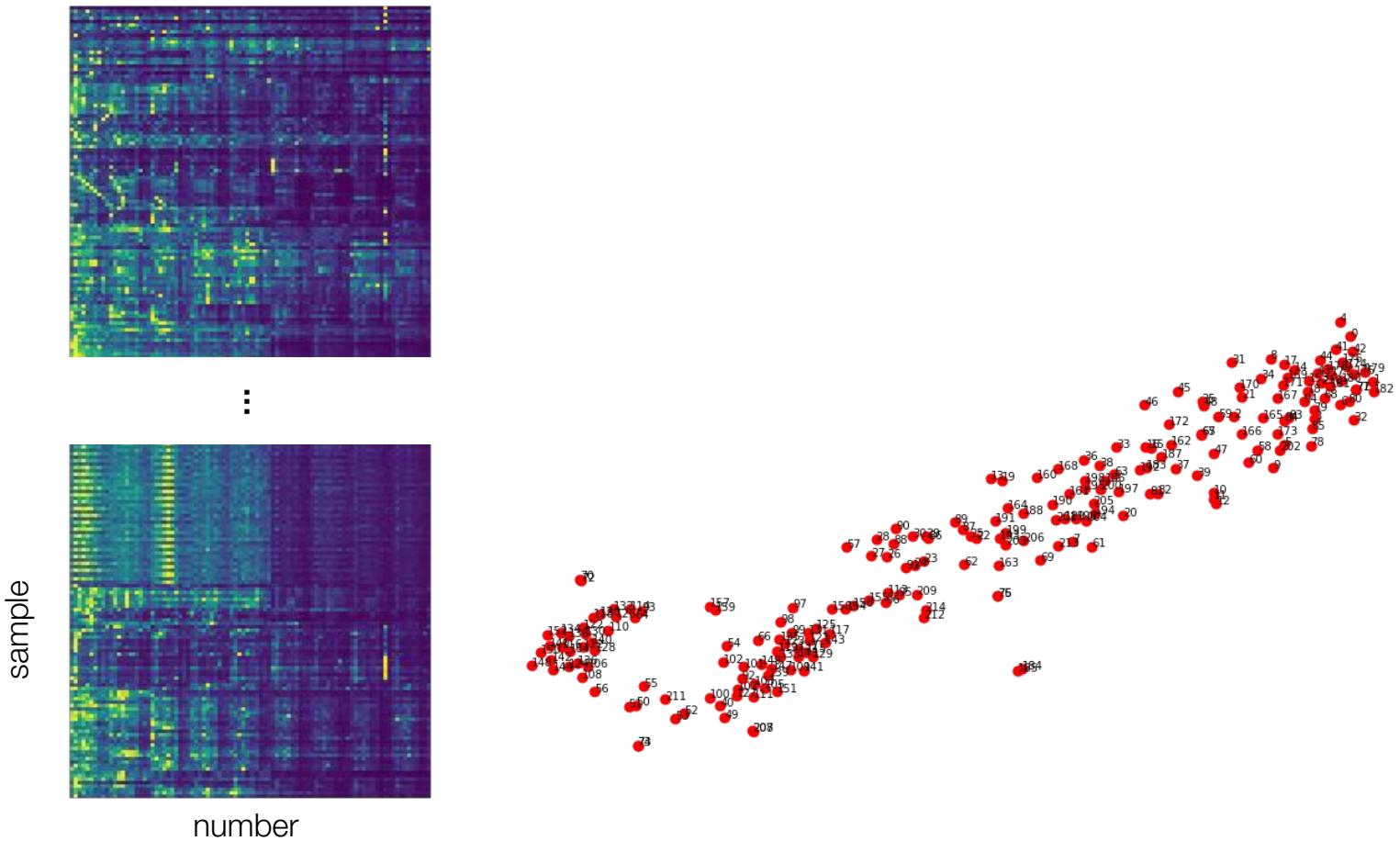
Cap. V.
Zodiaco sunt innotescere, quia et p[ro]pt[er] Zodiaco caro-
quatu[m] climatia declinationis. Secund[um] p[ro]p[ter] Zodiaco caro-
quatu[m] climatia declinationis. Secund[um] p[ro]p[ter] Zodiaco caro-
quatu[m] climatia declinationis. Duo ita et ceteri magis in Mod-
ulo admodum fideliter significare possunt alii et quoniam h[ab]ent p[er]fecta
affinitatem cum aliis fullitissimis poli finali transire diffinire Ecclipti-
cas. Coloris etiam imperfecti vocantur circuli H[ab]ent Aquariorum et etiam Za-
diaco etiam in quatu[m] iniquitate quadrantes ab aliis aquinoctiib[us] atque
fullinib[us] patiuntur, et p[re]cedentes diuiduntur p[re]cedentes equi-



Most similar



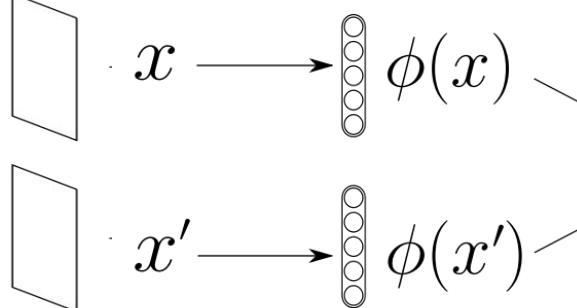
Modelling table pairs



II. Interpretable similarity models

Explaining similarity

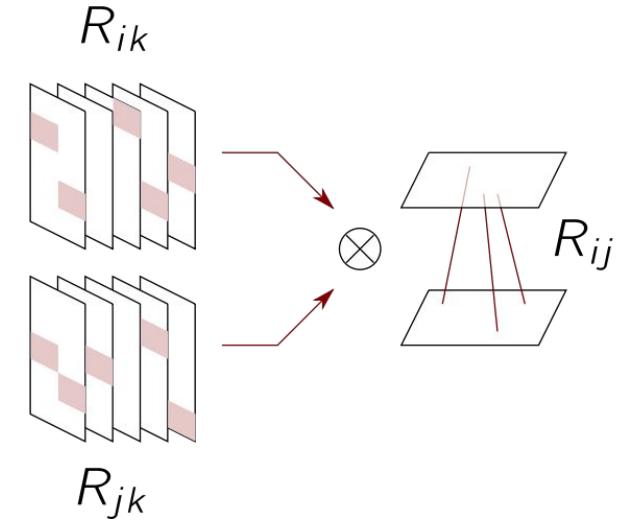
1. Extracting features



2. Compute similarity

$$y = \langle \phi(x), \phi(x') \rangle$$

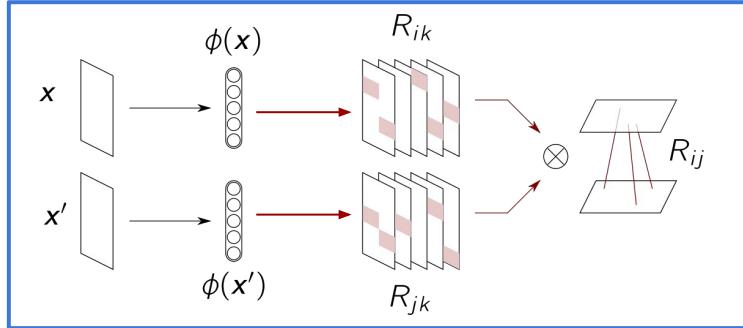
3. Propagate similarity scores



Propagation of similarity scores

Assuming dot-product similarity

$$y(\mathbf{x}, \mathbf{x}') = \sum_m y_m(\mathbf{x}, \mathbf{x}')$$



and mapping features using a *positively homogeneous function*, i.e.

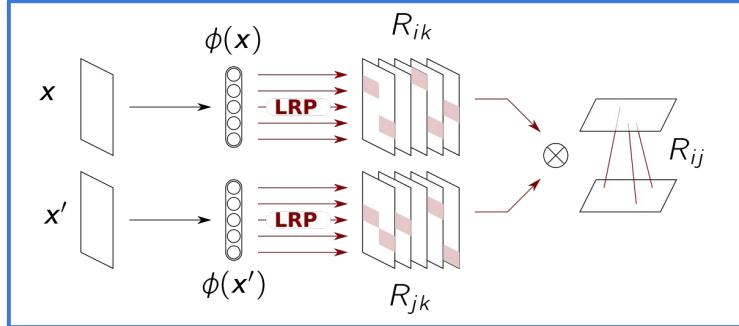
$$\forall t > 0 : \phi(t\mathbf{x}) = t\phi(\mathbf{x})$$

relevances $R_{ii'}$ can be factorised into two independent LRP branches for each feature m

Propagation of similarity scores

Assuming dot-product similarity

$$y(\mathbf{x}, \mathbf{x}') = \sum_m y_m(\mathbf{x}, \mathbf{x}')$$



and mapping features using a *positively homogeneous function*, i.e.

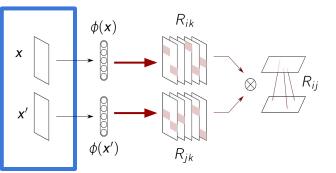
$$\forall t > 0 : \phi(t\mathbf{x}) = t\phi(\mathbf{x})$$

relevances $R_{ii'}$ can be factorised into two independent LRP branches for each feature m

$$R_{ii'} = \sum_m \underbrace{[\text{LRP}(\phi_m, \mathbf{x})]_i}_{R_{im}} \cdot \underbrace{[\text{LRP}(\phi_m, \mathbf{x}')]_{i'}}_{R_{i'm}}$$

Results

I. Explaining table similarity

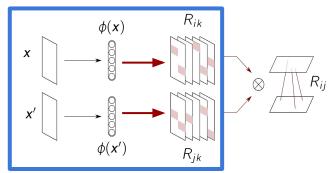


V		γ	
G.	M.	G.	M.
0	0	11	
0	22	12	
0	44	12	3
1	6	12	

	<i>V</i>	<i>γ</i>	II	III	IV	V	VI	VII	VIII	IX		
TABVL A ASCENSIONVM Obliquarum.												
	G.	M.	G.	M.	G.	M.	G.	M.	G.	M.		
1	0	0	11	18	27	42	74	91	46	135	50	
2	0	32	12	43	18	21	55	21	93	31	137	12
3	0	44	12	30	18	43	51	18	94	38	138	48
4	1	83	12	22	20	44	57	44	99	35	144	46
5	1	99	12	50	30	60	55	28	109	18	162	11
6	2	12	12	28	12	51	68	21	100	51	134	41
7	2	15	12	60	12	12	61	67	74	146	11	
8	3	18	12	24	12	12	66	12	71	146	11	
9	3	21	12	54	12	12	66	21	70	146	11	
10	3	24	12	42	12	12	66	53	104	47	149	10
11	3	33	12	16	12	12	54	53	104	35	149	58
12	4	33	12	42	12	12	66	53	104	47	149	10
13	4	49	12	32	12	12	66	39	104	35	149	58
14	5	52	12	32	12	12	66	39	104	35	149	58
15	5	55	12	32	12	12	66	39	104	35	149	58
16	6	58	12	32	12	12	74	31	105	51	149	58
17	6	62	12	47	12	14	74	31	105	51	149	58
18	6	64	12	47	12	14	74	31	105	51	149	58
19	7	68	12	47	12	14	74	31	105	51	149	58
20	7	71	12	47	12	14	74	31	105	51	149	58
21	7	73	12	54	12	14	74	31	105	51	149	58
22	8	76	12	54	12	14	74	31	105	51	149	58
23	8	80	12	32	47	45	81	35	105	51	149	44
24	8	84	12	47	45	81	12	12	105	51	149	44
25	9	87	12	47	48	51	81	20	106	20	145	44
26	9	91	12	47	51	54	81	20	106	20	145	44
27	9	94	12	47	51	54	81	20	106	20	145	44
28	9	97	12	47	51	54	81	20	106	20	145	44
29	10	100	12	23	42	58	81	20	106	20	145	36
30	10	104	12	47	51	58	81	20	106	20	145	36
31	11	107	12	47	51	58	81	20	106	20	145	36
32	11	111	12	47	51	58	81	20	106	20	145	36
33	11	115	12	47	51	58	81	20	106	20	145	36
34	11	119	12	47	51	58	81	20	106	20	145	36
35	11	123	12	47	51	58	81	20	106	20	145	36
36	12	126	12	47	51	58	81	20	106	20	145	36
37	12	130	12	47	51	58	81	20	106	20	145	36
38	12	134	12	47	51	58	81	20	106	20	145	36
39	12	138	12	47	51	58	81	20	106	20	145	36
40	12	142	12	47	51	58	81	20	106	20	145	36
41	12	146	12	47	51	58	81	20	106	20	145	36
42	12	150	12	47	51	58	81	20	106	20	145	36

V		γ	
G.	M.	G.	M.
0	0	11	
0	22	12	
0	44	12	3

I. Explaining table similarity



feature map explanations ϕ_m

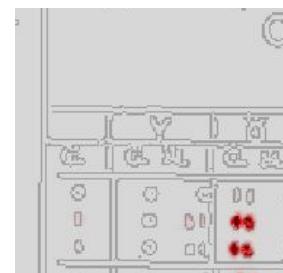
$m=11$

$m=12$

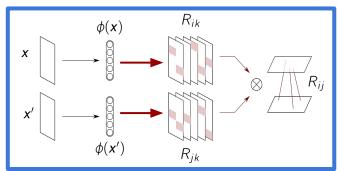
	V		ꝝ
	G.	M.	G.
	0	0	11
0	22	12	
0	44	12	
	1	6	12



	V		ꝝ
	G.	M.	G.
	0	0	11
0	22	12	
0	44	12	
	1	6	12



I. Explaining table similarity



feature map explanations ϕ_m

$m=11$

	V		λ
	G.	M.	G.
	0	0	11
	0	22	12
	0	44	12
	1	6	12

$m=12$

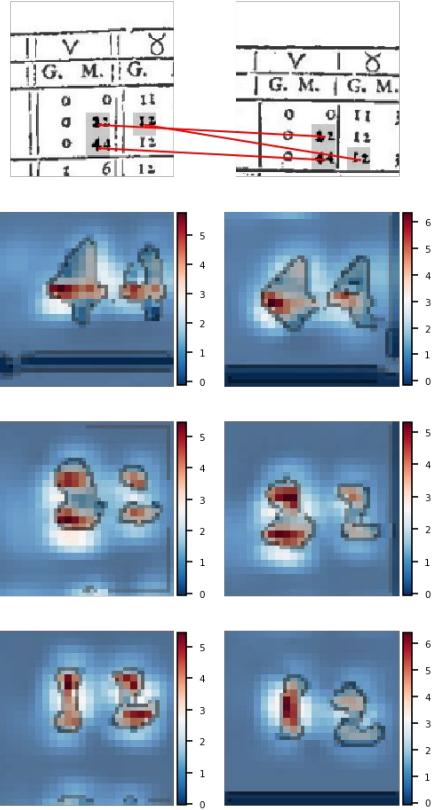


patch-level
explanations

	V		λ
	G.	M.	G.
	0	0	11
	0	22	12
	0	44	12
	1	6	12



pixel-level
explanations



Fine-grained explanations

LIBER II.				
Tabula declinationis Solis, per quēlibet gradum ecliptice.				
	Libra.	Scorpio	Sagitta.	Gemin.
grd.	gra. mi.	gra. mi.	gra. mi.	gra. mi.
1	0 24	11 51	20 25	29
2	0 48	12 12	20 37	28
3	1 12	12 33	20 49	27
4	1 36	12 53	21 0	26
5	1 59	13 13	21 11	25
6	2 23	13 33	21 22	24
7	2 47	13 53	21 32	23
8	3 11	14 13	21 42	22
9	3 35	14 32	21 52	21
10	3 58	14 51	22 0	20
II	4 22	15 10	22 9	19
12	4 45	15 28	22 17	18
13	5 9	15 47	22 25	17
14	5 32	16 5	22 32	16
15	5 55	16 23	22 39	15
16	6 19	16 40	22 46	14
17	6 42	16 57	22 52	13
18	7 5	17 14	22 57	12
19	7 28	17 31	23 3	11
20	7 50	17 47	23 7	10
21	8 13	18 3	23 12	9
22	8 35	18 19	23 15	8
23	8 58	18 34	23 19	7
24	9 20	18 49	23 22	6

LIBER II.				
Tabula declinationis Solis, per quēlibet gradum ecliptice.				
	Libra.	Scorpio	Sagitta.	Gemin.
grd.	gra. mi.	gra. mi.	gra. mi.	gra. mi.
1	0 24	11 51	20 25	29
2	0 48	12 12	20 37	28
3	1 12	12 33	20 49	27
4	1 36	12 53	21 0	26
5	1 59	13 13	21 11	25
6	2 23	13 33	21 22	24
7	2 47	13 53	21 32	23
8	3 11	14 13	21 42	22
9	3 35	14 32	21 52	21
10	3 58	14 51	22 0	20
II	4 22	15 10	22 9	19
12	4 45	15 28	22 17	18
13	5 9	15 47	22 25	17
14	5 32	16 5	22 32	16
15	5 55	16 23	22 39	15
16	6 19	16 40	22 46	14
17	6 42	16 57	22 52	13
18	7 5	17 14	22 57	12
19	7 28	17 31	23 3	11
20	7 50	17 47	23 7	10
21	8 13	18 3	23 12	9
22	8 35	18 19	23 15	8
23	8 58	18 34	23 19	7
24	9 20	18 49	23 22	6

3	I 3
3	3 3
3	5 3

2	I 2
2	3 3
2	5 3

2 1
2 1
2 2

2 0
2 1
2 1

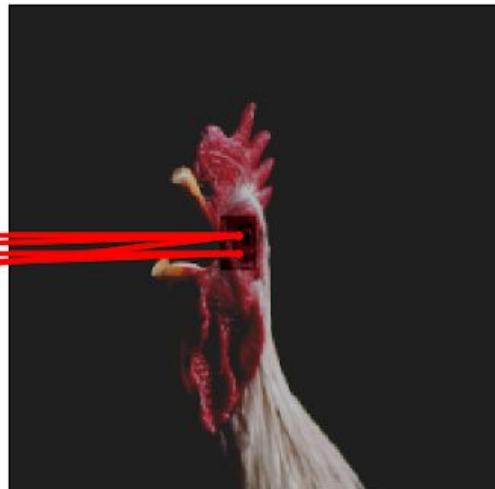
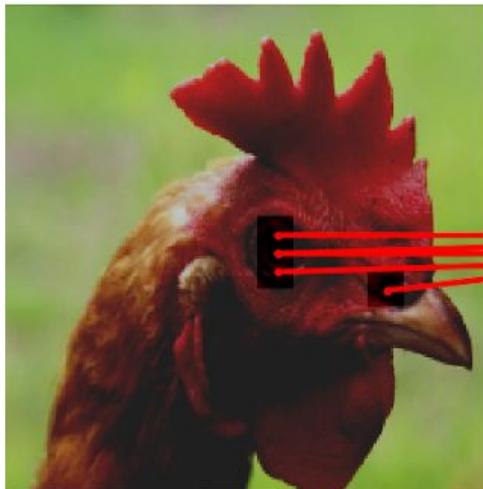
II. Explaining natural image pairs

Why are these two images similar?



II. Explaining natural image pairs

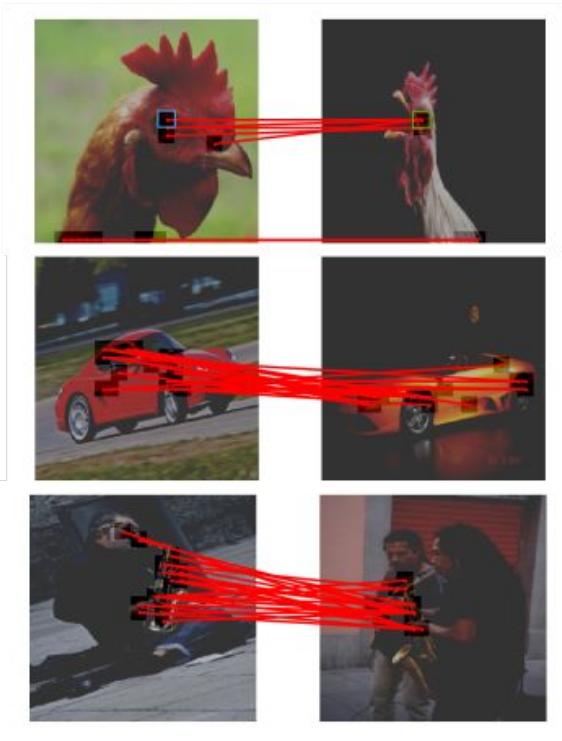
Why are these two images similar?



II. Explaining natural image pairs

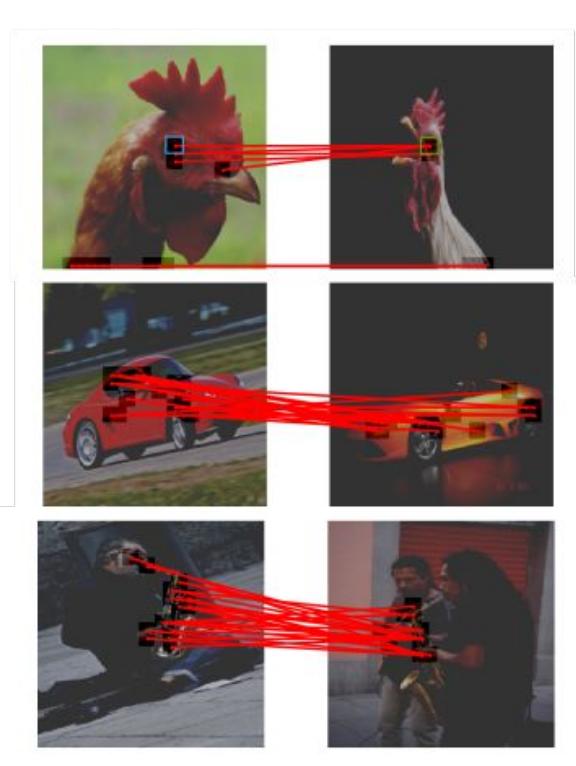
patch-level explanations

pixel-level explanations

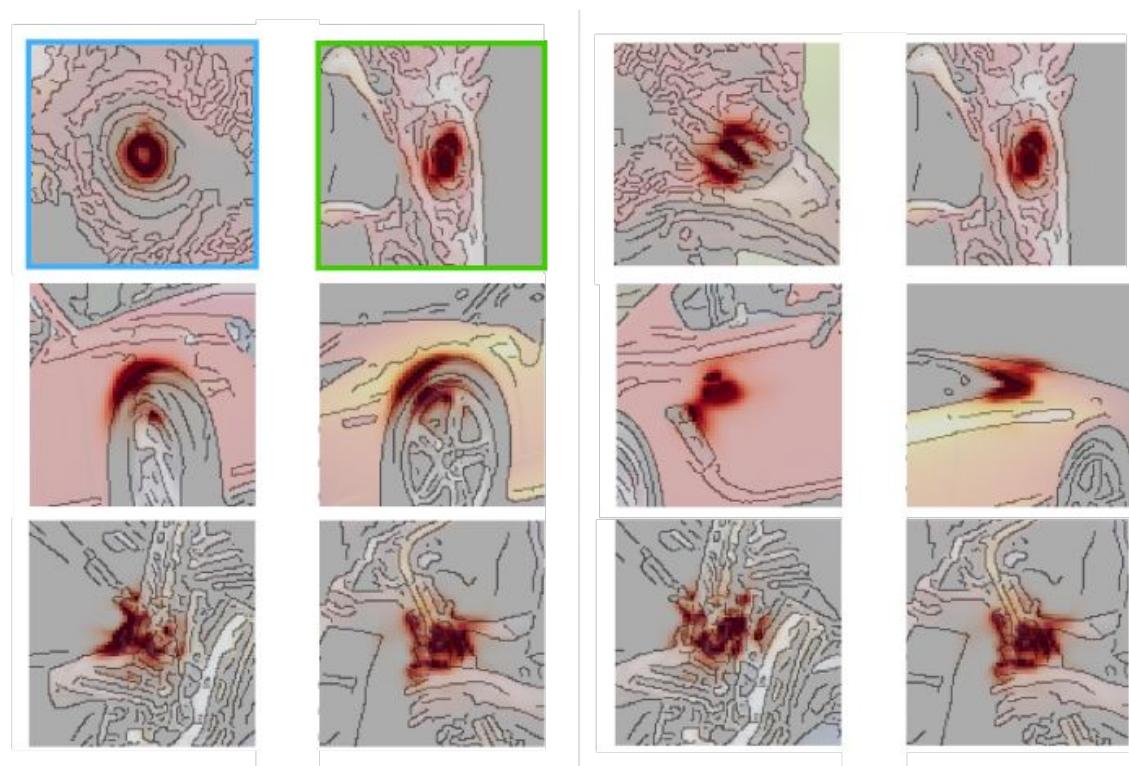


II. Explaining natural image pairs

patch-level explanations

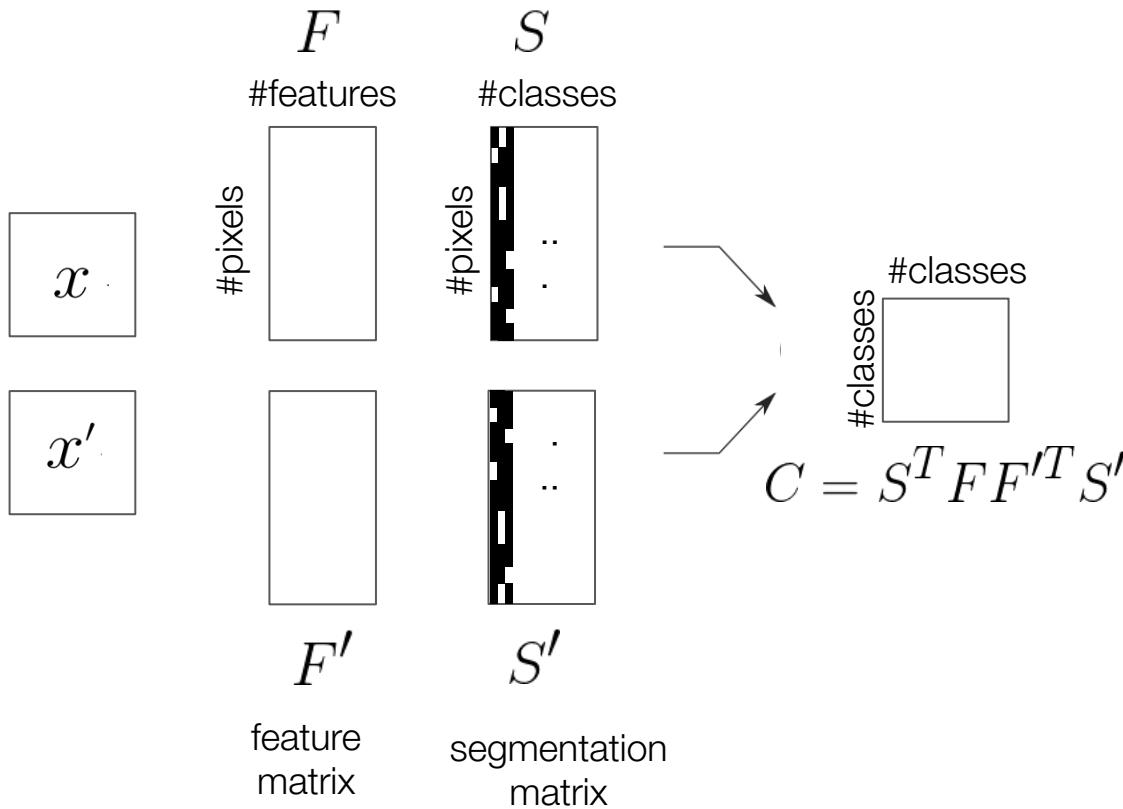


pixel-level explanations



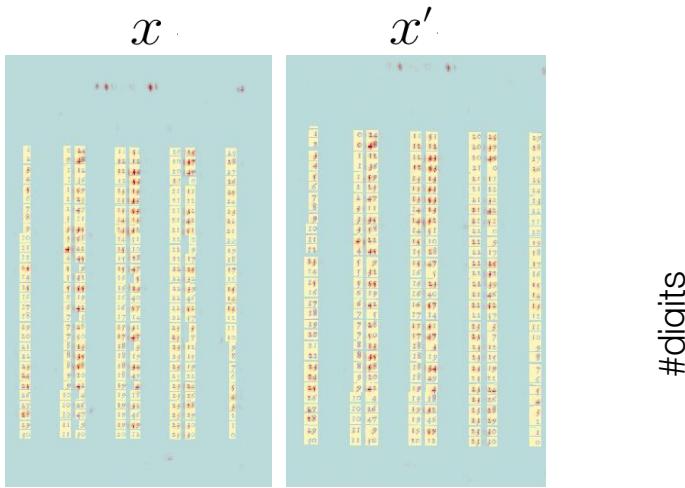
III. Evaluation

How to evaluate explained similarity?

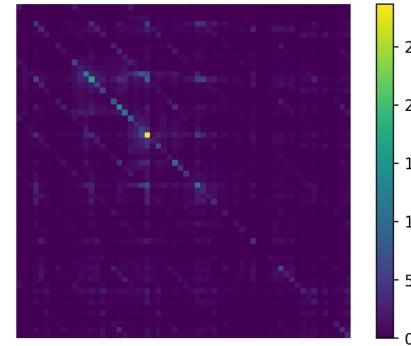


How to evaluate explained similarity? (work in progress)

I. Tables



#digits

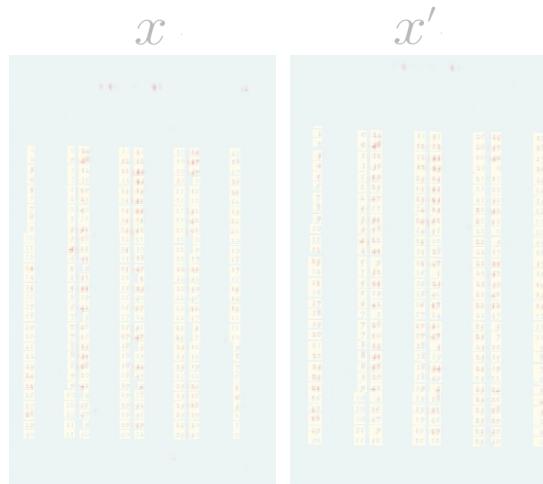


Evaluation on fully annotated tables (number and bounding box)

- In-vs-outside bounding box flow: 80.02%
- Number-level box flow: 61.42%
(chance level 2%,
flow between n_numbers+1 classes!)

How to evaluate explained similarity? (work in progress)

I. Tables



Evaluation on fully annotated tables (number and bounding box)

- In-vs-outside bounding box flow: 80.02%
- Number-level box flow: 61.42%
(chance level 2%,
flow between n_numbers+1 classes!)

II. Natural images



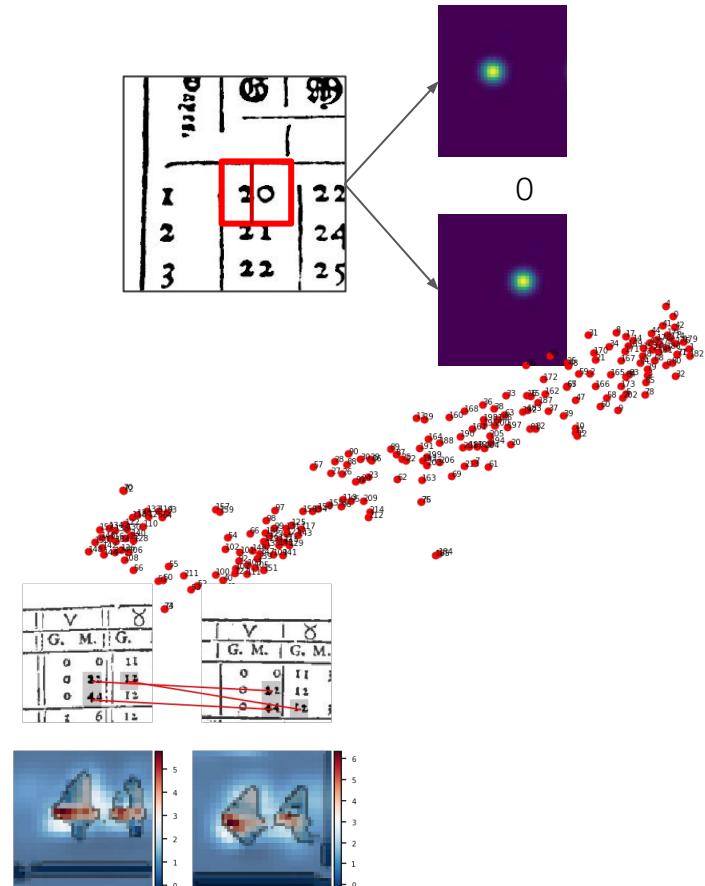
UCF Sports Action Dataset

Evaluating background-vs-ROI flow

Is similarity grounded in annotated regions or
mostly determined by background?

Summary

1. A bag-of-features approach is proposed which learns a representation of numerical tables
2. A small set of features is learned requiring only few annotated data points
3. This table network is an example for explainable deep similarity models (new!)
4. Flow across segmentation masks is used to evaluate explained similarity



G.	G.	M.	G.	M.	G.	M.	G.	M.	G.	M.
0	180	0	220	30	261	-2	297	48	325	26
1	181	20	221	51	262	21	298	53	326	11
2	182	41	223	13	263	40	299	57	326	56
3	184	1	224	34	264	59	301	1	327	40
4	185	22	225	56	266	18	302	4	328	24
5	186	43	227	18	267	37	303	6	329	7
6	188	3	228	40	268	55	304	8	329	49
7	189	24	230	2	270	12	305	9	330	30
8	190	45	231	23	271	29	306	9	331	12
9	192	6	232	45	272	46	307	9	331	53
10	193	27	234	-6	274	2	308	8	332	34
11	194	47	235	28	275	18	309	6	333	15
12	196	18	236	49	276	34	310	4	333	55
13	197	29	238	11	277	49	311	1	334	34
14	198	50	239	32	279	4	311	57	335	13
15	200	11	240	-53	280	18	312	52	335	51
16	201	32	242	15	281	32	313	47	336	29
17	202	53	243	36	282	45	314	41	337	6
18	204	14	244	58	283	58	315	35	337	43
19	205	35	246	19	285	10	316	28	338	20
20	206	56	247	40	286	22	317	20	338	56
21	208	17	249	1	287	33	318	11	339	32
22	209	38	250	22	288	44	319	2	340	8
23	211	0	251	42	289	54	319	52	340	43
24	212	21	253	3	291	3	320	42	341	19
25	213	43	254	23	292	12	321	31	341	54
26	215	4	255	43	293	20	322	19	342	3
27	216	25	257	3	294	28	323	6	343	2
28	217	47	258	23	295	35	323	53	343	35
29	219	-8	259	43	296	42	324	40	344	31

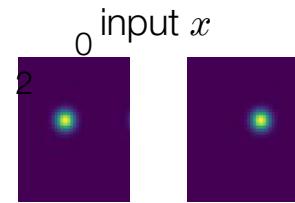
Sparse annotations

Tabula declinationis Solis, cuiuslibet gradus Ecliptice maxima declinatione 23. graduum, 28. minutorum suffutata.

N	Libra.	Scorpi.	Sagit.
odi.	Aries.	Taurus.	Gemini
G.	G.M. 2.	G.M. 2.	G.M. 2. G.
0	0 0 0	11 29 5	20 20 15
1	0 2; 5;	11 50 6	20 22 17
2	0 47 4; 12	10 56	10 5 7 28
3	1 11 39 12	31 34 20 15	55 27
4	1 5 30 12	51 59	10 58 20 26
5	1 59 20 13	12 22 21	9 21 29
6	1 2; 8; 1;	32 12 21	19 59 24
7	2 40 5; 13	51 58 21	30 19 23
8	2 10 37 14	21 10 21	40 5 22
9	3 34 18 14	10 48 21	46 19 28
10	3 57 5; 14	49 51 21	53 19 20
11	3 23 28 15	8 40 12	7 6 19
12	4 44 5; 15	27 13 22	15 17 18
	Virgo.	Leo.	Cancer.
	Pisces.	Aquar.	Capric.

Represent compositions

0	5	10
1	20	22
2	21	24
3	22	25



$$\phi_m(x) = h_i(x)h_j(x * \delta)$$

Embed and compare

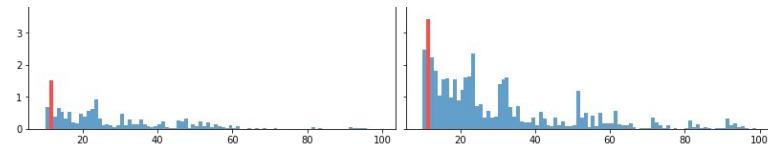
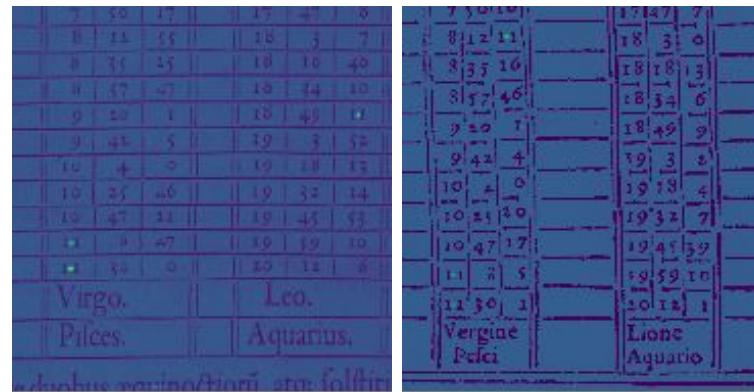


Table comparison

Table activation $h_{i=1,j=1}$

	Libra.	Aries.	Scorpio	Taurus.	Sagitta.	Gemin.	
	gra.	mi.	gra.	mi.	gra.	mi.	gra.
1	0	24	11	51	20	25	29
2	0	48	12	11	20	37	28
3	1	12	12	33	20	49	27
4	1	36	12	53	21	0	26
5	1	59	13	13	21	11	25
6	2	23	13	33	21	22	24
7	2	47	13	53	21	32	23
8	3	11	14	13	21	42	22
9	3	35	14	32	21	51	21
10	3	58	14	51	22	0	20
11	4	22	15	10	22	9	19
12	4	45	15	28	22	17	18
13	5	9	15	47	22	25	17

Table histogram $s_m = \sum_i \phi_m(x_i)$

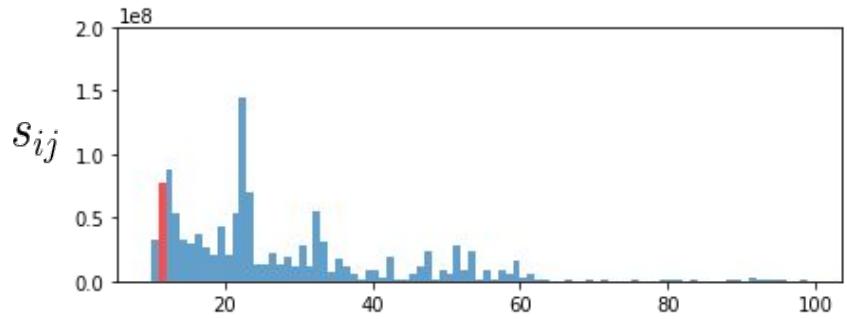


Table Embeddings - annotated data

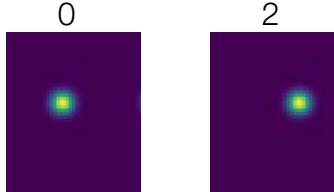
P A R S
*Tabula declinationis Solis, cuiuslibet
gradus Eclipticae maxima declinatione
23. graduum, 28. minutorum
supputata.*

Nodi	Libra.	Scorpi.	Sagit.	
Aries.	Taurus.	Gemini		
G.	G.M. 2.	G.M. 2.	G.M. 2.	G.
0	0 0 0	11 29 5	20 20 15	30
1	0 23 5;	11 50 6	20 22 17	29
2	0 47 46	12 10 56	10 5 7	28
3	1 11 39	12 31 34	20 15 55	27
4	1 15 36	12 51 59	20 58 20	26
5	1 59 20	12 22 21	9 21 25	
6	2 23 8;	32 12 21	19 59 24	
7	2 40 54;	51 58 21	30 19 23	
8	2 10 37 14	21 30 21	40 5 22	
9	3 34 18 14	30 48 21	46 19 21	
10	3 57 5 14	49 51 21	53 19 20	
11	3 21 28 15	8 40 22	7 6 19	
12	4 44 5 15	27 13 22	15 17 18	
Virgo.	Leo.	Cancer.		
Pisces.	Aquar.	Capric.		

Representing compositions

	0	60	90
0			
1			
2			
3			

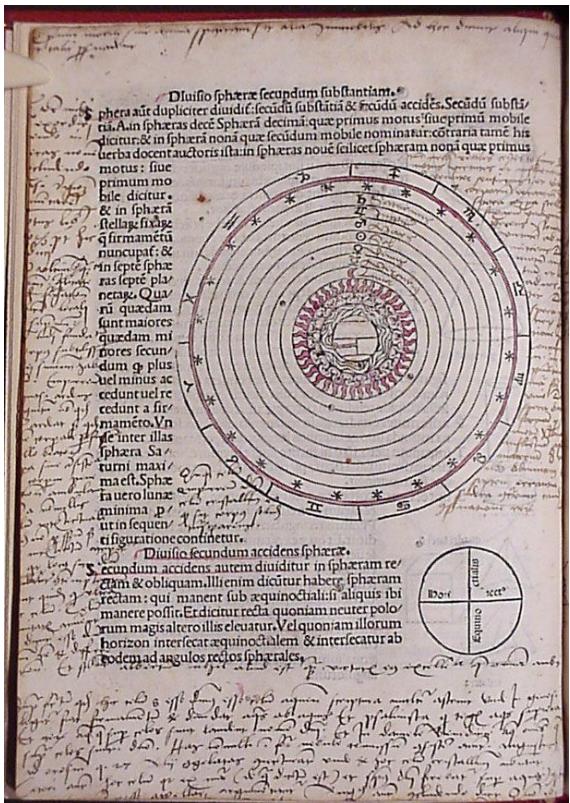
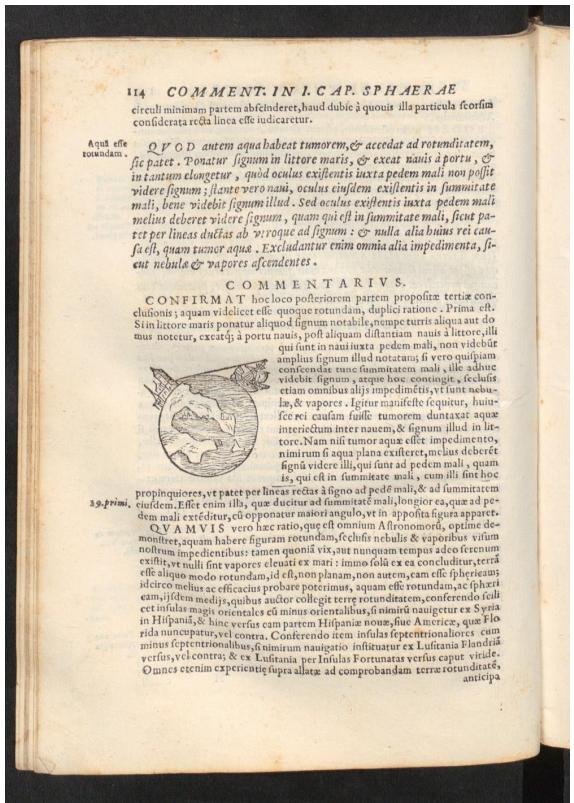
input x



Masking context

	0	60	90
0			
1			
2			
3			

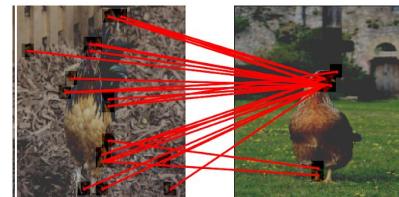
Sphaera corpus: learning to represent tables



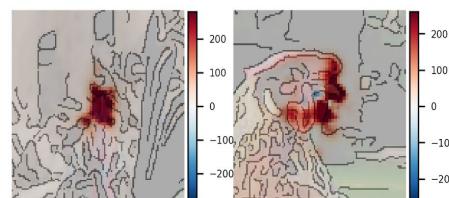
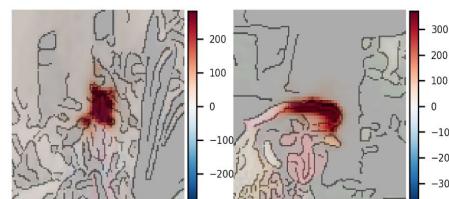
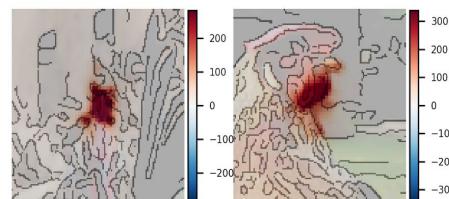
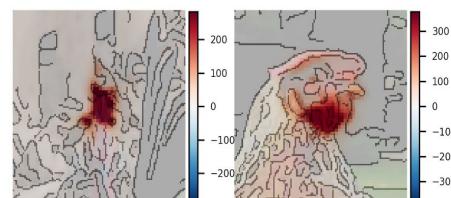
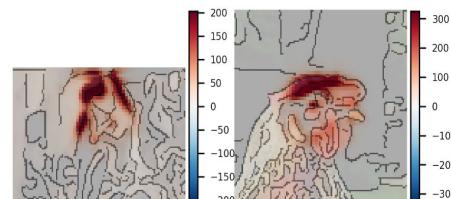
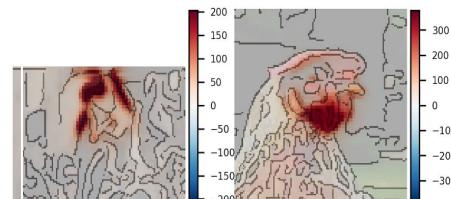
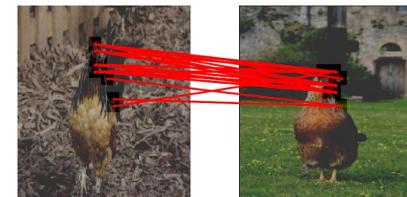
Julius.				Augustus.			
Soi ausf. gang	Soi Nite- gang	Tags Leng.	Nacht Leng.	Soi ausf. gang	Soi Nite- gang	Tags Leng.	Nacht Leng.
st. m.	st. m.	st. m.	st. m.	st. m.	st. m.	st. m.	st. m.
1 3. 52 8. 8 16	7 44	1 4.25 7.37 15	10 8 50				
2 3. 53 8. 7 16	7 46	2 4.27 7.33 15	6 8 54				
3 3. 53 8. 7 16	7 46	3 4.28 7.32 15	4 8 56				
4 3. 54 8. 6 16	7 48	4 4.29 7.31 15	2 8 58				
5 3. 55 8. 5 16	10 50	5 4.30 7.30 15	0 9 0				
6 3. 55 8. 5 16	10 50	6 4.36 7.29 14	8 9 2				
7 3. 56 8. 4 16	8 52	7 4.34 7.26 14	52 9 8				
8 3. 57 8. 3 16	6 54	8 4.35 7.25 14	50 9 10				
9 3. 57 8. 3 16	6 54	9 4.37 7.23 14	46 9 14				
10 3. 58 8. 2 16	4 56	10 4.38 7.22 14	44 9 16				
II 3. 58 8. 2 16	4 56	11 4.40 7.20 14	40 9 20				
12 3. 59 8. 1 16	2 58	12 4.42 7.18 14	36 9 24				
13 4. 0 8. 0 16	0 0	13 4.43 7.17 14	34 9 26				
14 4. 1 7. 19 15	58 2	14 4.45 7.15 14	30 9 30				
15 4. 2 7. 58 15	56 4	15 4.47 7.13 14	26 9 34				
16 4. 3 7. 57 15	54 6	16 4.49 7.11 14	22 9 38				
17 4. 4 7. 56 15	52 8	17 4.51 7.9 14	18 9 42				
18 4. 5 7. 55 15	50 10	18 4.53 7.7 14	14 9 46				
19 4. 6 7. 53 15	48 14	19 4.54 7.6 14	12 9 48				
20 4. 8 7. 52 15	44 16	20 4.56 7.4 14	8 9 52				
21 4. 9 7. 51 15	42 18	21 4.58 7.2 14	4 9 56				
22 4.11 7.49 15	38 22	22 5. 0 7.0 14	0 10 0				
23 4.12 7.48 15	36 24	23 5. 2 6.58 13	56 10 4				
24 4.13 7.47 15	34 26	24 5. 4 6.56 13	52 10 8				
25 4.15 7.45 15	30 30	25 5. 5 6.55 13	50 10 10				
26 4.16 7.44 15	28 32	26 5. 7 6.53 13	46 10 14				
27 4.17 7.43 15	26 34	27 5. 9 6.51 13	42 10 18				
28 4.19 7.41 15	22 38	28 5.11 6.49 13	38 10 22				
29 4.21 7.39 15	18 42	29 5.13 6.47 13	34 10 26				
30 4.22 7.38 15	16 44	30 5.16 6.44 13	28 10 32				
31 4.24 7.36 15	12 48	31 5.17 6.43 13	26 10 34				

II. Explaining high-vs-low-level features

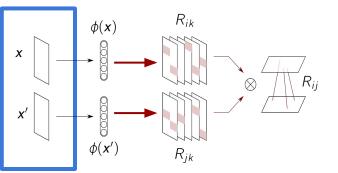
vgg17



vgg31



I. Explaining table similarity



	V		γ
G.	M.	G.	
0	0	11	
0	22	12	
0	44	12	
1	6	12	

384 Comment. in III. Cap. Sphere
TABVLA ASCENSIONVM
Obliquarum.

	V	G.	M.	G.
1	0	0	11	
2	0	22	12	
3	0	44	12	
4	1	6	12	
5	2	12	12	
6	3	18	12	
7	4	24	12	
8	5	30	12	
9	6	36	12	
10	7	42	12	
11	8	48	12	
12	9	54	12	
13	10	60	12	
14	11	66	12	
15	12	72	12	
16	13	78	12	
17	14	84	12	
18	15	90	12	
19	16	96	12	
20	17	102	12	
21	18	108	12	
22	19	114	12	
23	20	120	12	
24	21	126	12	
25	22	132	12	
26	23	138	12	
27	24	144	12	
28	25	150	12	
29	26	156	12	
30	27	162	12	

384 Comment. in 3. Cap. Sphere
TABVLA ASCENSIONVM
Obliquarum.

	V	G.	M.	G.
1	0	0	11	
2	0	22	12	
3	0	44	12	
4	1	6	12	
5	2	12	12	
6	3	18	12	
7	4	24	12	
8	5	30	12	
9	6	36	12	
10	7	42	12	
11	8	48	12	
12	9	54	12	
13	10	60	12	
14	11	66	12	
15	12	72	12	
16	13	78	12	
17	14	84	12	
18	15	90	12	
19	16	96	12	
20	17	102	12	
21	18	108	12	
22	19	114	12	
23	20	120	12	
24	21	126	12	
25	22	132	12	
26	23	138	12	
27	24	144	12	
28	25	150	12	
29	26	156	12	
30	27	162	12	

	V		γ
G.	M.	G.	M.
0	0	11	
0	22	12	
0	44	12	