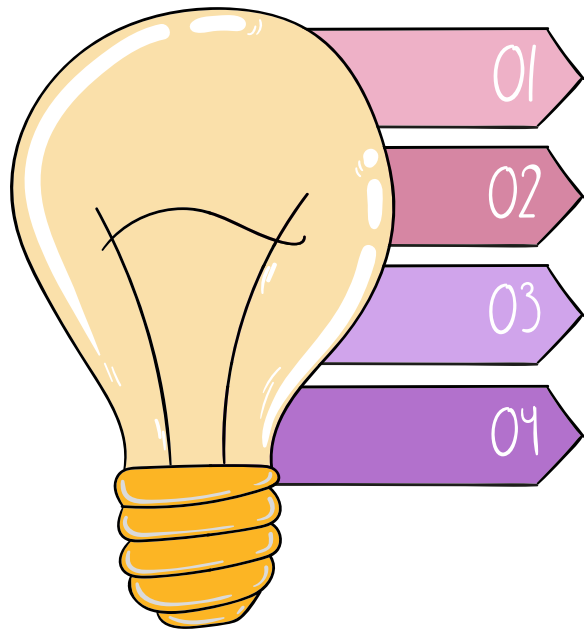


Methods for detecting relevant elements of charts and generating their description

Katarzyna Lorenc
Karolina Seweryn
Supervisor: Anna Wróblewska, PhD

Presentation plan

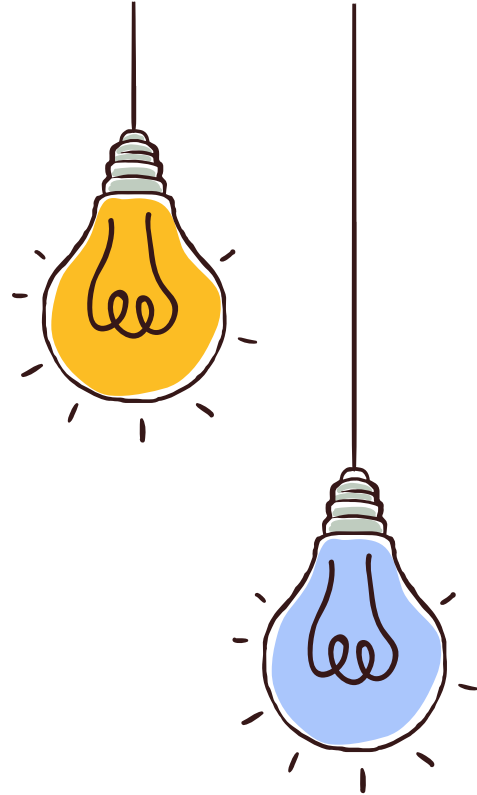


Datasets

Classification

Object detector

Text Generator





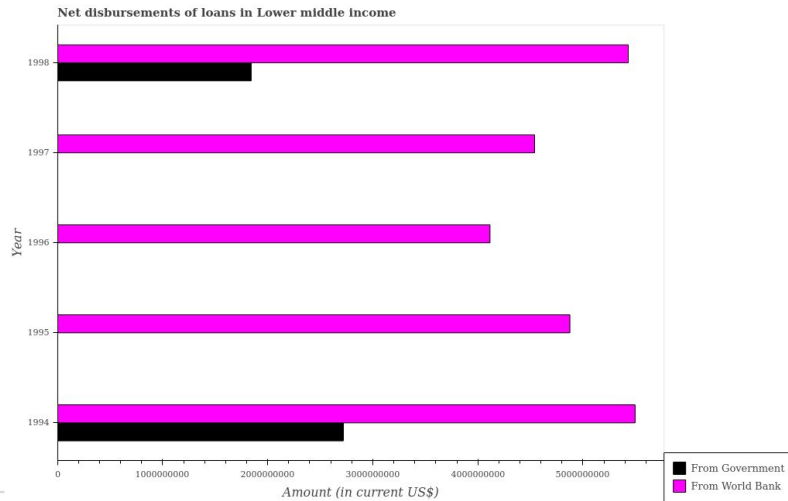
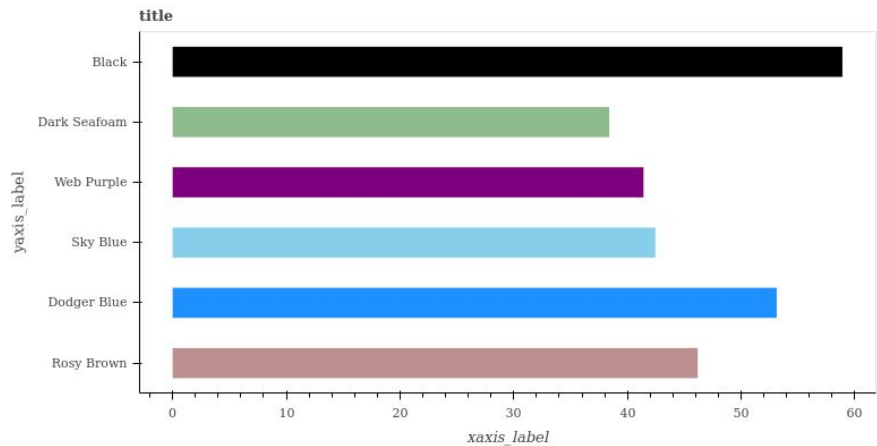
Datasets

The properties of analysed datasets

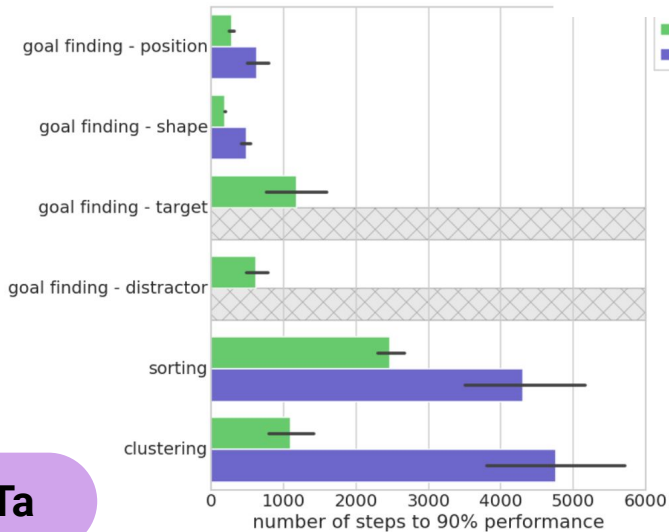


	FigureQA	PlotQA	ChaTa
Real values	✗	✓	✓
Real plots	✗	✗	✓
Number of figures	140 000	224 377	7 170
Data	Synthetic	Semi-Synthetic	Real

FigureQA

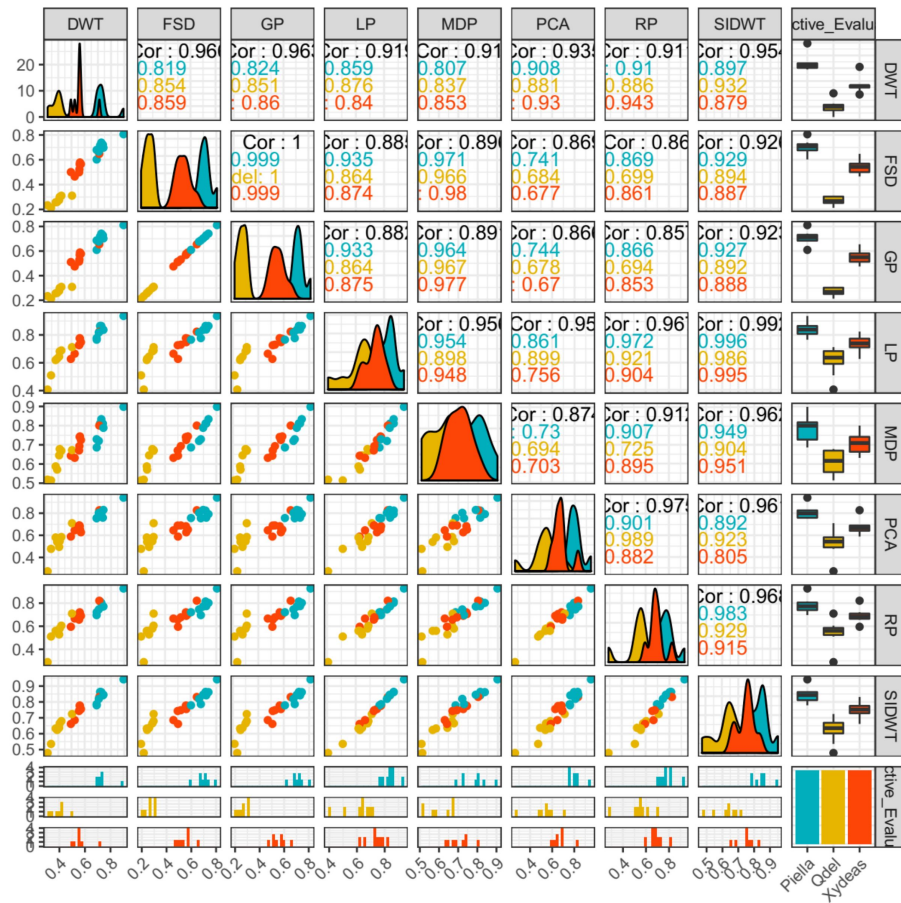
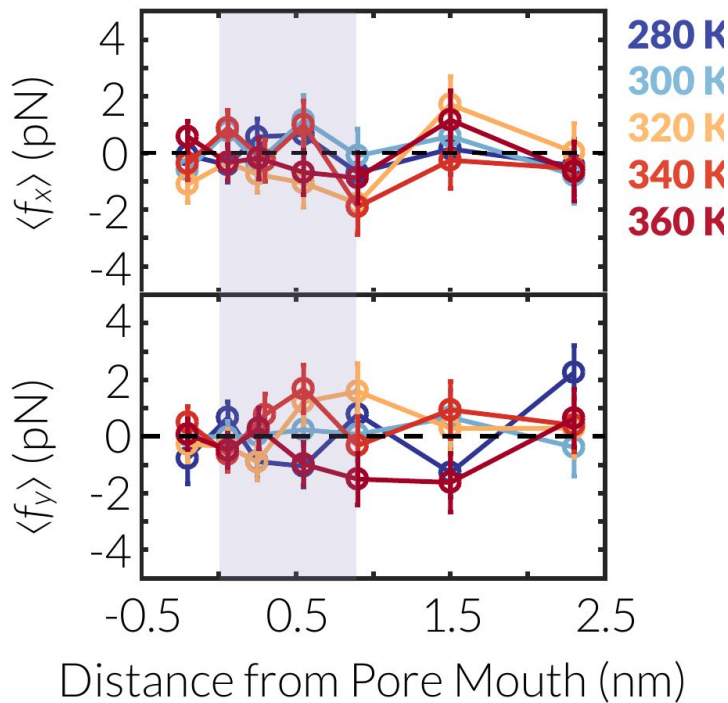


PlotQA



ChaTa

ChaTa - difficult examples



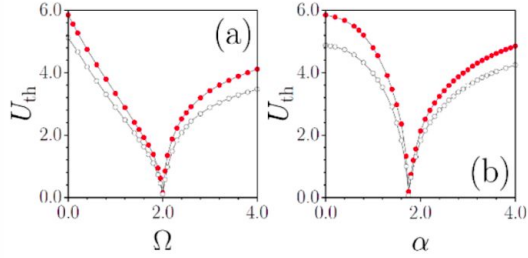


Fig. 3. The threshold norm, above which 2D solitons with the single-peak ($\Omega > 2\alpha^2$) and striped ($\Omega < 2\alpha^2$) structure exist, vs. Ω , at $\alpha = 1$ (a), and vs. α , at $\Omega = 6$ (b). Curves with solid and open dots correspond, severally, to $g_s = g_c = 1$ and $g_s = 1.2$, $g_c = 0.8$.

III. 3D system with 2D Rashba spin-orbit coupling

The 3D version of Eq. (1), where $\Delta_3 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$, is taken with the 2D Rashba SOC, i.e., $\alpha_y = \alpha_x \equiv \alpha$ and $\mathbf{h}^{[x]} = \hat{\mathbf{y}}$, $\mathbf{h}^{[y]} = -\hat{\mathbf{x}}$, corresponding to Hamiltonian $H_{\text{soc}} = \alpha(\hat{k}_x \sigma_y - \hat{k}_y \sigma_x)$, where $\hat{k}_{x,y}$ are the components of the momentum operator. The respective linear dispersion relation for in-plane radial component of the wave vector, $k_r \equiv (k_x^2 + k_y^2)^{1/2}$, and the orthogonal component, k_z , is

$$\mu_{\pm} = (k_r^2 + k_z^2) / 2 \pm (\alpha^2 k_r^2 + \Omega^2 / 4)^{1/2}. \quad (5)$$

Dependencies of the threshold norm U_{th} on Ω and α , as obtained from the numerical solution of Eq. (1), are presented in Fig. 3 (see the curves with solid dots corresponding to $g_s = g_c = 1$). Note that $U_{\text{th}} \rightarrow U_{\text{TOWIDES}}$ in both limits of $\Omega \rightarrow 0$ (when SOC can be gauged away from Eq. (1) with the Manakov nonlinearity [35,36]) and $\Omega \rightarrow \infty$ [making the ψ_1 component vanishingly small and reducing Eq. (1) to the single GPE for ψ_2].

Figure

Figure
Caption

Description
in text

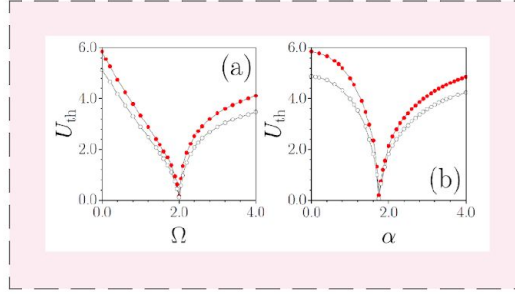
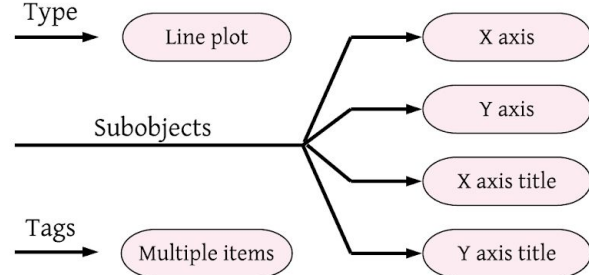


Fig. 3. The threshold norm, above which 2D solitons with the single-peak ($\Omega > 2\alpha^2$) and striped ($\Omega < 2\alpha^2$) structure exist, vs. Ω , at $\alpha = 1$ (a), and vs. α , at $\Omega = 6$ (b). Curves with solid and open dots correspond, severally, to $g_s = g_c = 1$ and $g_s = 1.2$, $g_c = 0.8$.

Dependencies of the threshold norm U_{th} on Ω and α , as obtained from the numerical solution of Eq. (1), are presented in Fig. 3 (see the curves with solid dots corresponding to $g_s = g_c = 1$).



ChaTa+

Figure Types	ChaTa	ChaTa+	Figure Types
Line plot	4,027	974	Line plot
Dot plot	1,655	338	Dot-Line plot
Other plot	1,502	–	–
H barplot	37	26	H barplot
V barplot	425	302	V barplot
Box plot	360	–	–
Pie chart	11	–	–
Sum	7,171	1,640	

Figure Elements	
Y axis	1 618
X axis	1 567
Chart description	1 407
Legend	853
X axis title	552
Y axis title	265
Title	167
Reference in text	64

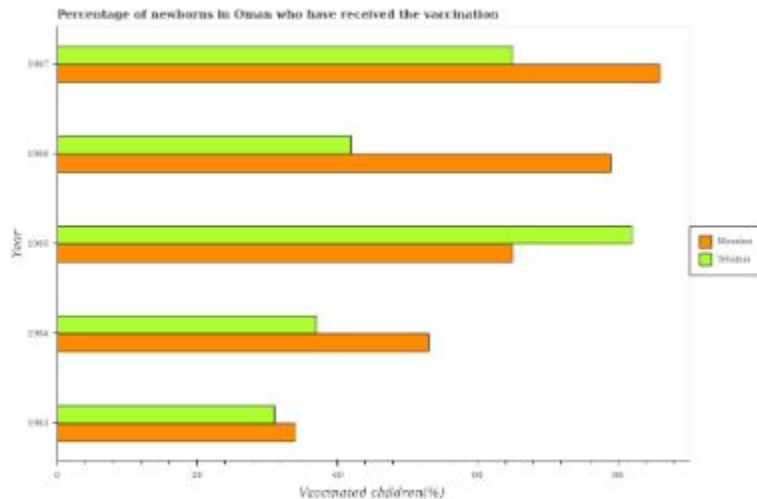
Examples of question-answer pairs from PlotQA

Questions	Answers	Sentences
Are all the bars in the graph horizontal?	Yes	All bars in the graph are horizontal.
What is the title of the graph?	"Number of servers"	The title of the graph is " Number of servers ".
Does "Haiti" appear as one of the legend labels in the graph?	No	"Haiti" doesn't appear as one of the legend labels in the graph.
How many legend labels are there?	4	There are 4 legend labels.



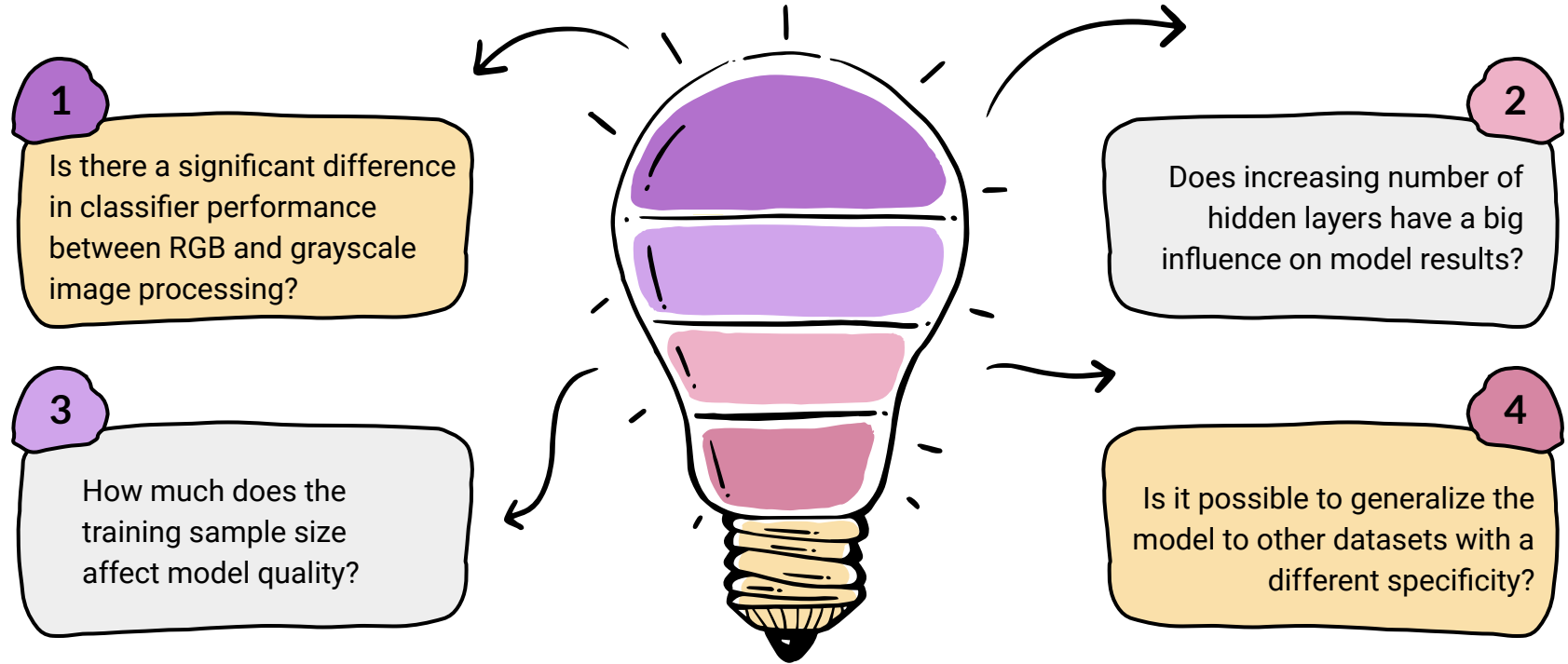
Classification

Differences between charts and natural-scene images

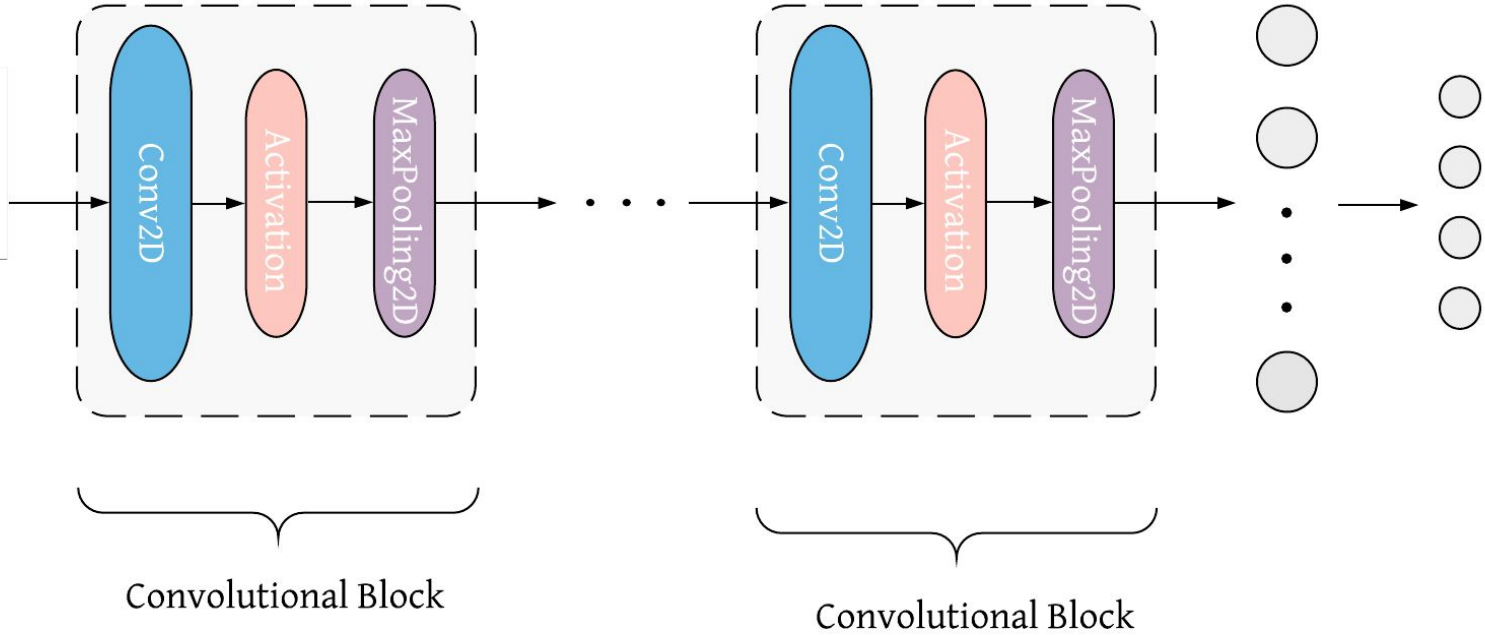
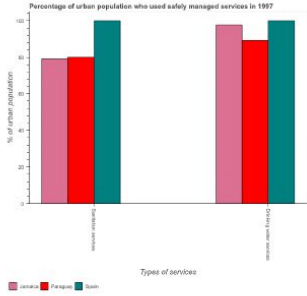


- (1) charts are covered with large areas of the same colour,
- (2) texts and their position in figures are crucial for interpretation,
- (3) the relations between chart elements should be interpreted precisely (e.g. color of barplot and corresponding legend box).

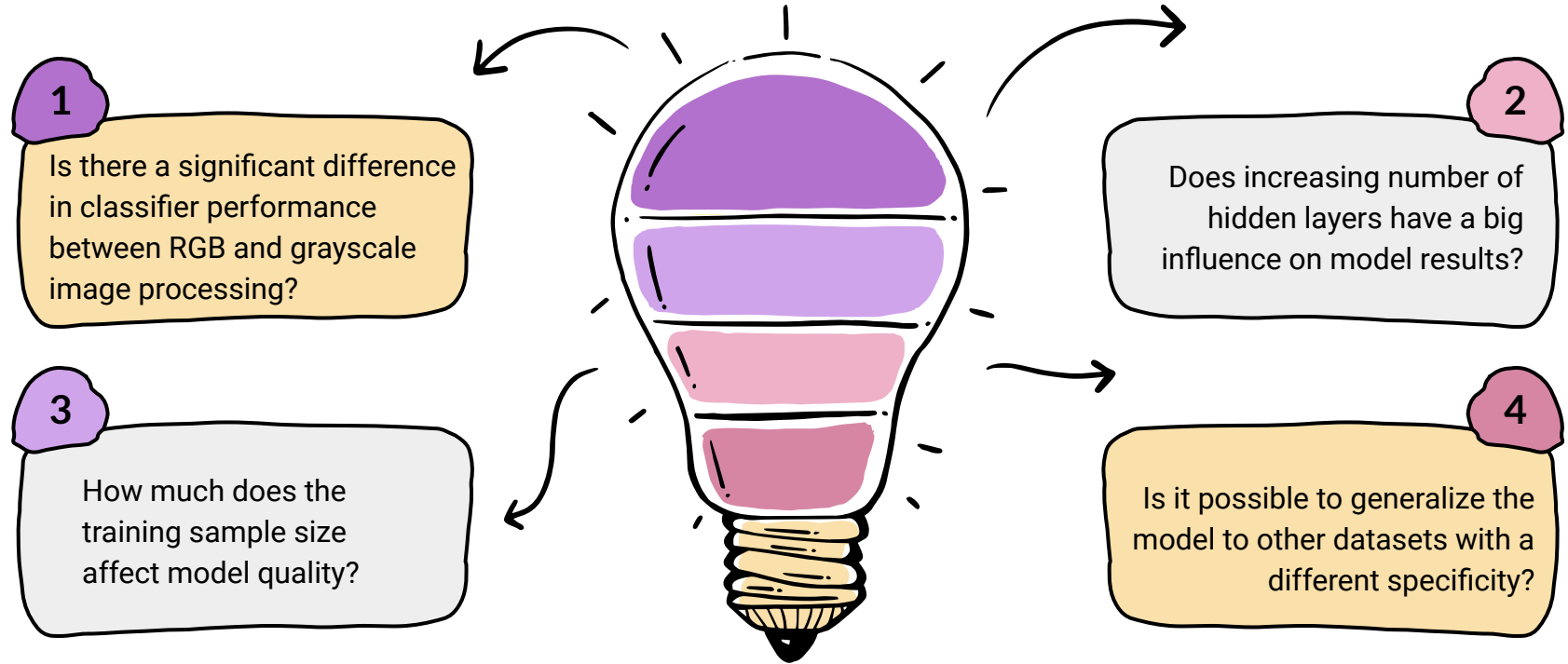
Research Questions



Classifier architecture



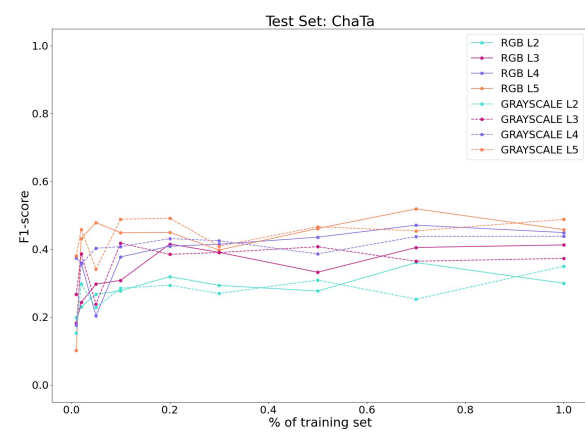
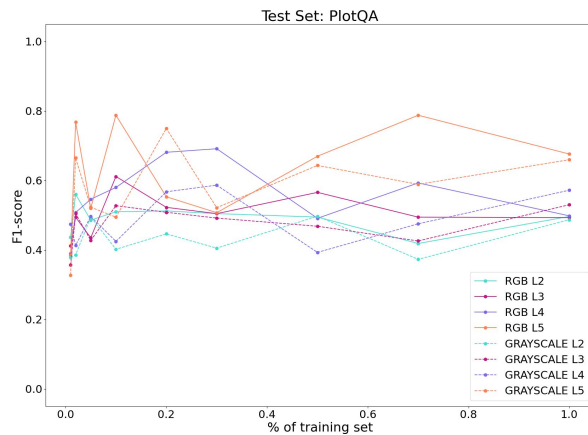
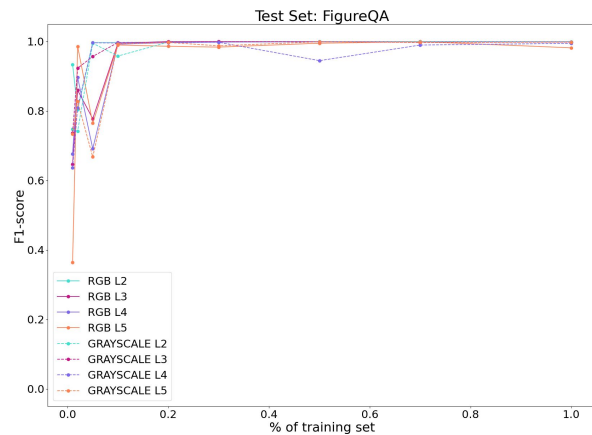
Research Questions



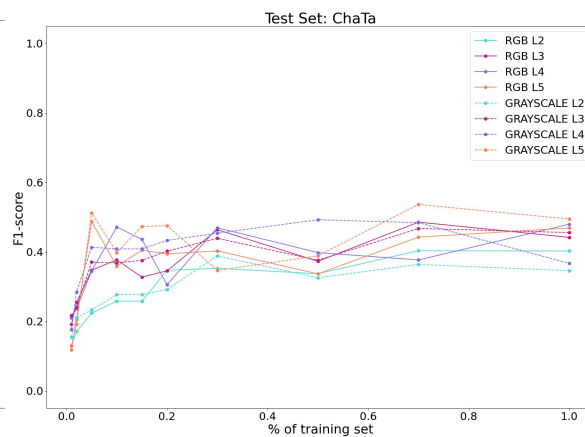
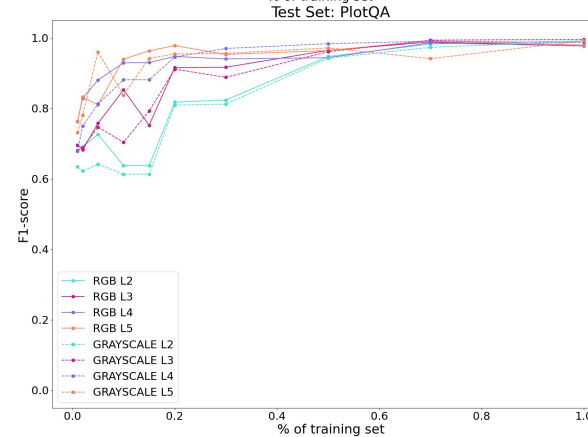
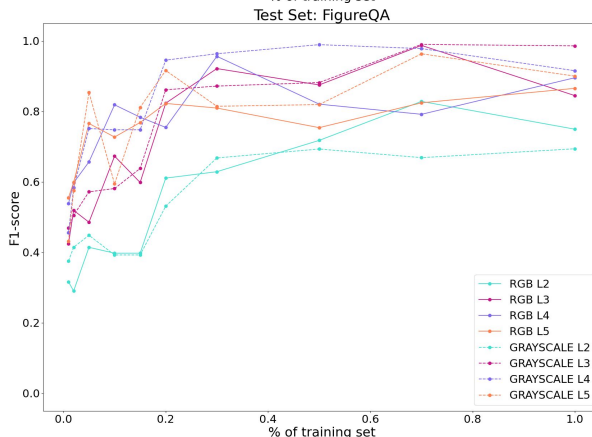
Training set

Classification results

FigureQA



PlotQA



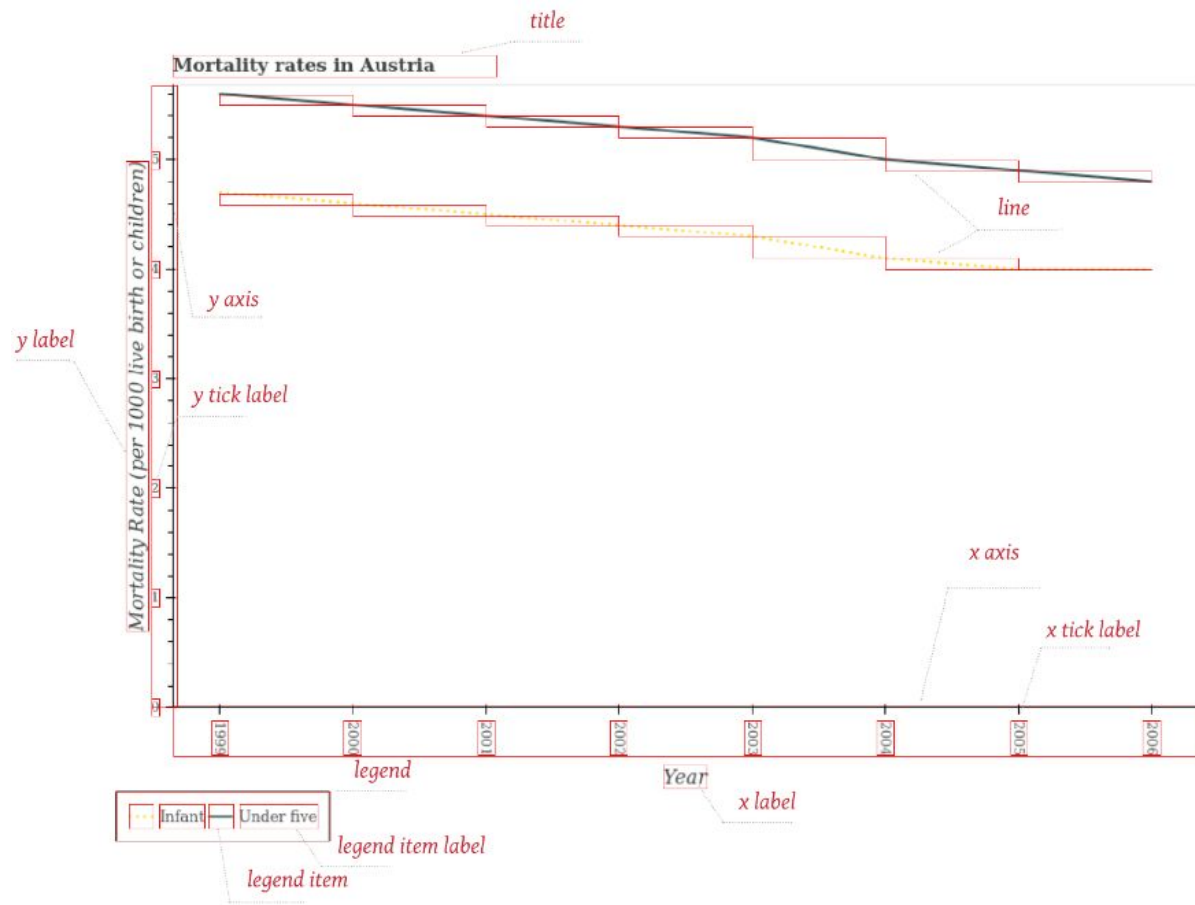
Test set	Model	Accuracy	F1-score	Precision	Recall	TOPSIS*
FigureQA	L3 grayscale	0.985	0.985	0.986	0.985	2
	L5 grayscale	0.903	0.899	0.927	0.903	1
	SVM grayscale	0.879	0.871	0.907	0.879	3
	SVM RGB	0.769	0.727	0.823	0.769	5
	ResNet	0.957	0.957	0.957	0.957	4
PlotQA	L3 grayscale	0.996	0.995	0.995	0.995	2
	L5 grayscale	0.994	0.992	0.992	0.992	1
	SVM grayscale	0.993	0.989	0.989	0.989	3
	SVM RGB	0.996	0.995	0.995	0.995	5
	ResNet	0.906	0.848	0.909	0.860	4
ChaTa+	L3 grayscale	0.620	0.455	0.506	0.586	2
	L5 grayscale	0.676	0.495	0.503	0.571	1
	SVM grayscale	0.548	0.396	0.440	0.607	3
	SVM RGB	0.358	0.286	0.385	0.519	5
	ResNet	0.487	0.406	0.423	0.528	4



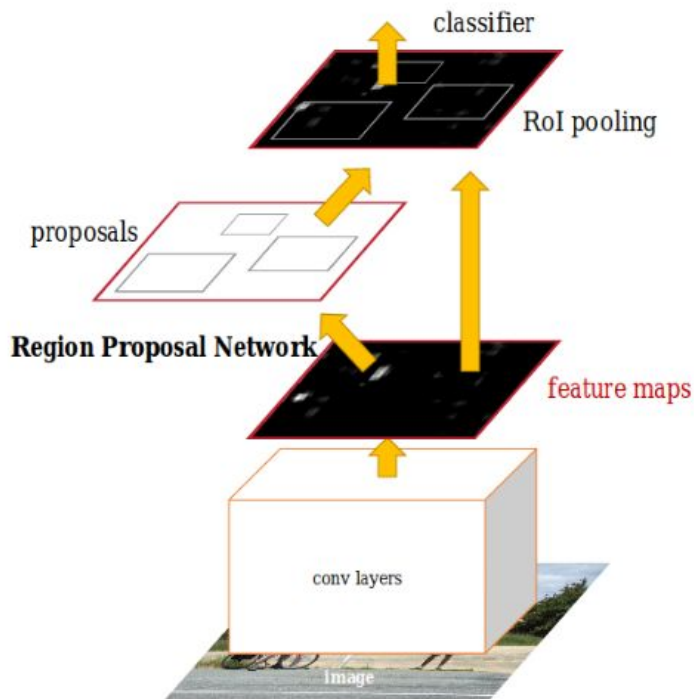
Object detection

Object Detection

- 13 classes of objects



Models - Faster R-CNN



- Improved version of region-based CNN (RCNN)
- Two parts: Fast R-CNN and Region Proposal Network

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*),$$

where:

i is an index of anchor,

p_i is a predicted probability that anchor i is an object,

p_i^* is ground-truth label, so it is equal to 1 if anchor is assigned correctly and 0 otherwise,

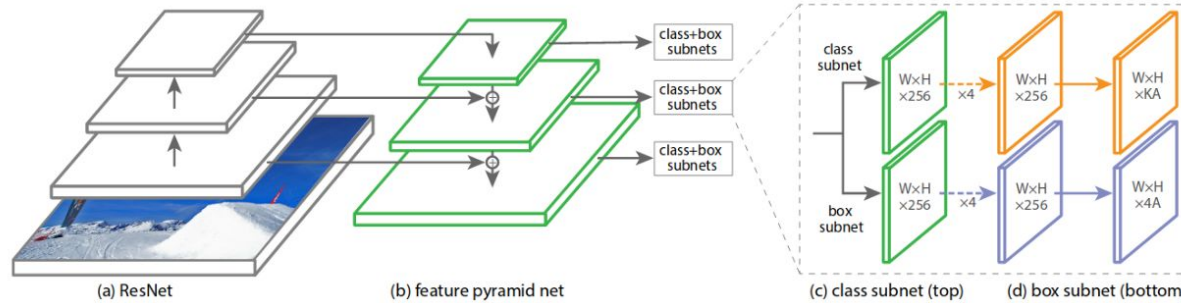
t_i means coordinates of predicted bounding box,

t_i^* is real bounding box associated with anchor (if the anchor exists),

L_{cls} is a classification loss,

L_{reg} is a regression loss.

Models - RetinaNet



Focal loss:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t),$$

where $\gamma \geq 0$.

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}.$$

Importance of input image size

- Faster R-CNN
- Training dataset: PlotQA

AP	200 x 150	400 x 250	600 x 400	800 x 500	1000 x 650
bar	40.55	63.05	73.01	79.41	82.53
dot line	1.26	21.52	42.66	55.17	60.99
legend	69.52	87.96	94.22	92.22	94.72
line	22.71	32.73	41.10	48.12	53.51
title	79.29	86.32	84.73	88.23	91.01
x label	23.71	85.30	89.50	93.08	94.91
x tick label	0.00	7.64	36.97	42.68	46.52
y label	3.90	84.50	96.25	98.36	98.31
y tick label	0.10	6.05	25.90	38.89	46.72
x axis	78.84	91.40	92.46	92.12	94.97
y axis	95.03	99.02	98.74	97.30	96.67
legend element	0.02	0.00	37.43	72.52	77.44
legend element label	0.13	1.08	57.10	78.55	83.91

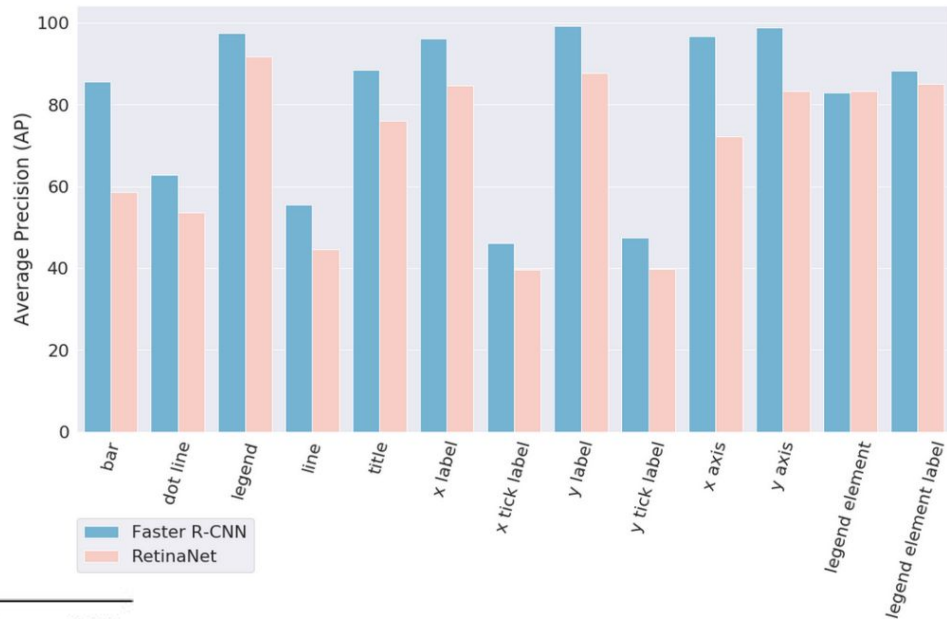
	Image Size	mAP	mAP50	mAP75	mAPs	mAPm	mAPI
Faster R-CNN	200 x 150	30.93	43.11	36.18	1.20	32.97	62.60
	400 x 250	51.28	62.33	55.32	19.69	63.65	86.37
	600 x 400	66.93	80.63	73.48	45.06	75.31	86.19
	800x500	75.13	87.65	83.66	58.82	79.56	90.34
	1000x650	78.63	88.87	86.12	64.28	81.96	91.83

The bigger image
the better model
performance



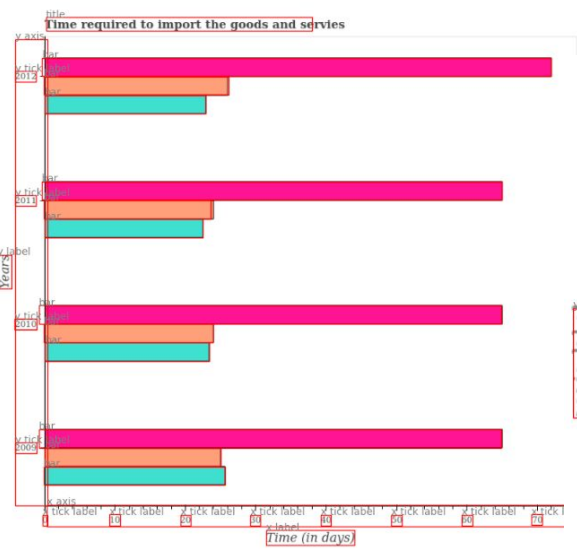
Chosen model

- Faster R-CNN
- Training dataset: 50% of PlotQA
- Image size: 700x1100



Model	Test set	mAP	mAP50	mAP75	mAPs	mAPm	mAPl
Faster R-CNN	FigureQA	33.33	58.69	30.13	19.51	40.37	54.01
	PlotQA	80.45	89.07	86.71	66.62	83.35	93.48
RetinaNet	FigureQA	29.32	56.17	23.40	21.85	35.11	47.83
	PlotQA	69.26	86.15	79.57	59.22	71.17	65.84

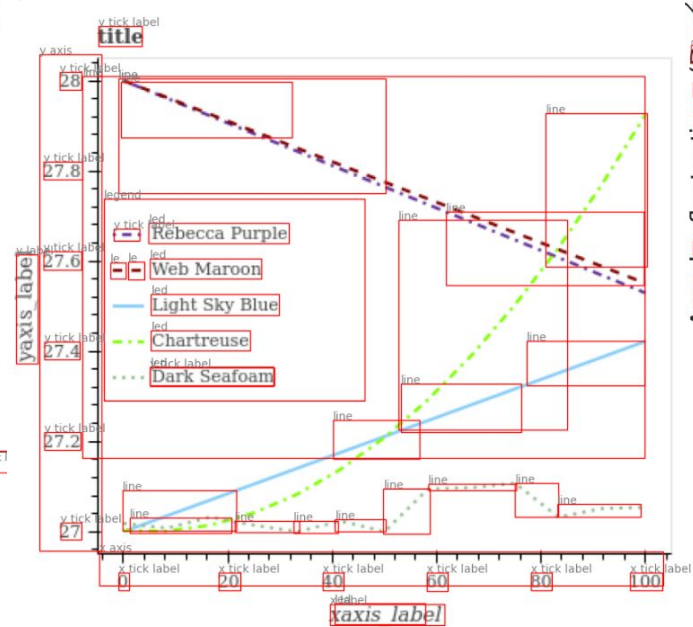
1



Almost infallible on PlotQA

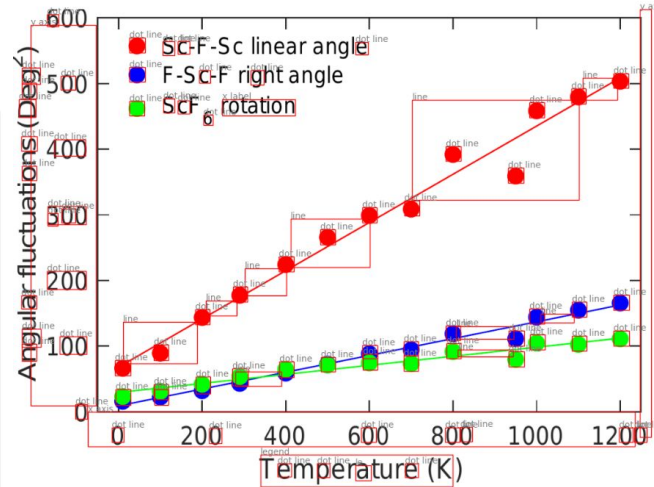
Examples and Findings

2



Difficulties when legend is in the center of the chart

3



Two y axes

Some letters are classified as dots

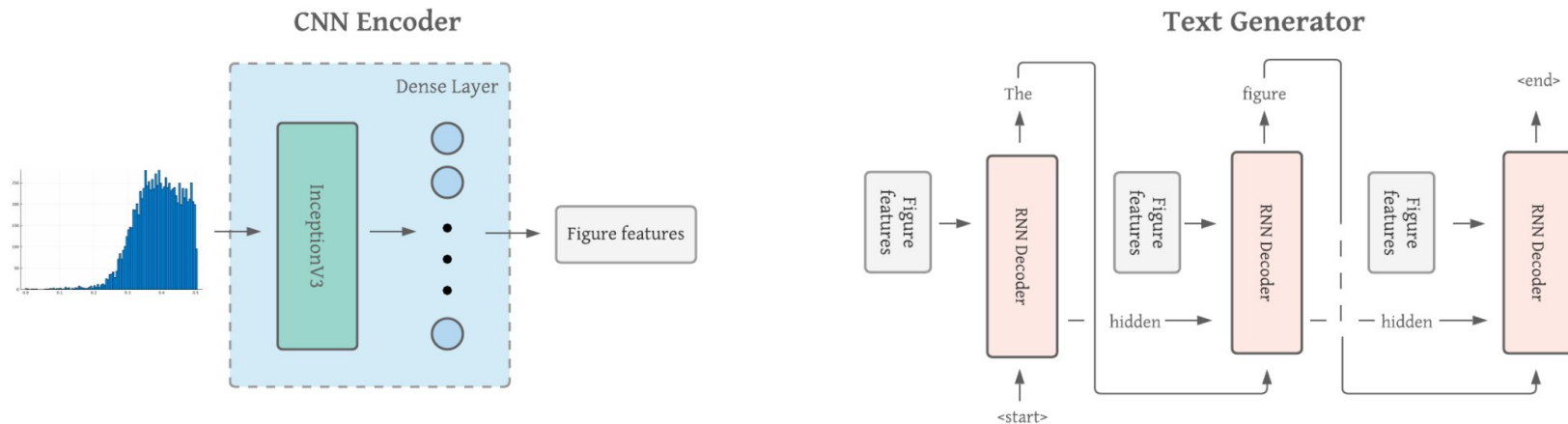


Text generation

Example of references and gold standard

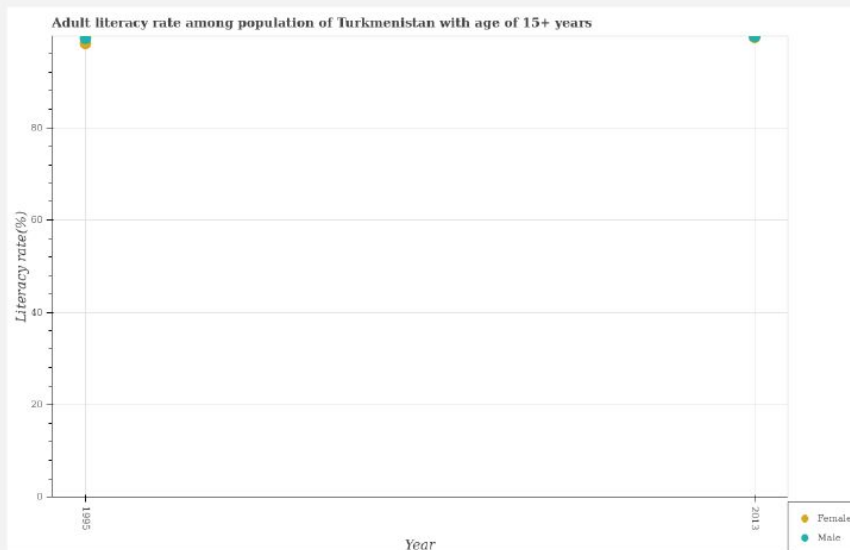
References	{'The figure presents the dependence of Cost(% of GNI per capita) on Years.', 'Chart type is line plot.', 'Legend labels are stacked vertically.', 'The title of the graph is Cost of business start-up procedures.', 'Montenegro doesn't appear as one of the legend labels in the graph.', 'The label or title of the X-axis is Years.', 'The label or title of the Y-axis is Cost(% of GNI per capita).'} }
Gold Standard	'The label or title of the X-axis is Years. Legend labels are stacked vertically. The title of the graph is Cost of business start-up procedures.'

Text generator I



	$BLEU_1$	$BLEU_2$	$BLEU_3$	$BLEU_4$	ROUGE-L
Model BA	37	24.70	15.55	9.41	31.05

Prediction of generator I for the observation from PlotQA.



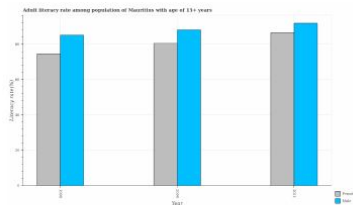
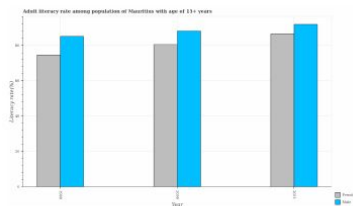
Real caption

The figure presents the dependence of Literacy rate(%) on Year. The label or title of the Y-axis is "Literacy rate(%)". There are 2 legend labels. The title of the graph is "Adult literacy rate among population of Turkmenistan with age of 15+ years". Legend labels are stacked vertically.

Prediction

The difference between the highest and the lowest export or 1975 attending of belgium is 0.

Text generator II



Slot type	Row	Slot value
Chart type	1	Vertical bar plot



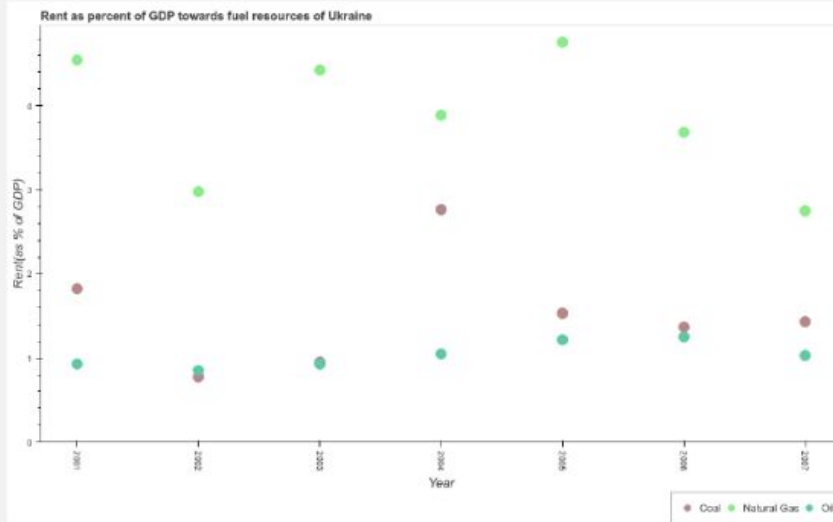
Slot type	Row	Slot value
Number of bars	2	6
Number of dot lines	3	0
Number of x tick label	4	3
X title	5	Year
...
Chart title	n	Adult literacy rate ...



The figure presents the dependence of Adult literacy

	BLEU ₁	BLEU ₂	BLEU ₃	BLEU ₄	ROUGLE-L
T2T	40	36.29	32.59	29.05	32.23

Prediction of generator II for the observation from PlotQA.



Real caption

Chart type is dot line plot . There are 3 different coloured dotlines . `` Slovak Republic does nt appear as one of the legend labels in the graph . The figure presents the dependence of Rent (as % of GDP) on Year .

Prediction

Chart type is dot line plot . The label or title of the X-axis is `` Year . There are 3 different coloured dotlines . The title of the graph is `` Percentage of repeaters in all grades of primary education in Equatorial Guinea .

Example



Text Generation

Object
Detection

Classification

The figure presents the dependence of Amount of exports (% of total goods exports) on Year . The label or title of the Y-axis is “ Amount of electricity produced (kWh) . **Chart type is dot line plot . There are 3 different coloured dotlines .**

[('image_id', 2519, 1), ('chart type', 'dot line plot', 2), ('legend element description', 'Natural gas sources', 3), ..., ('number of legend element', 3, 18), ('number of legend element description', 3, 19), ('max value of y', 40.0, 20)]

