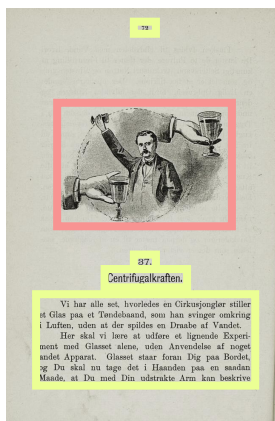


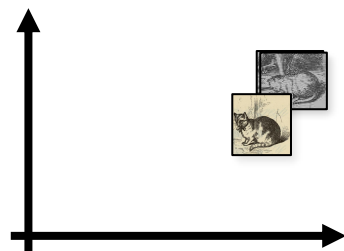
DH-LAB Bildesøk:

Nasjonalbibliotekets prototype på en bildesøkapplikasjon

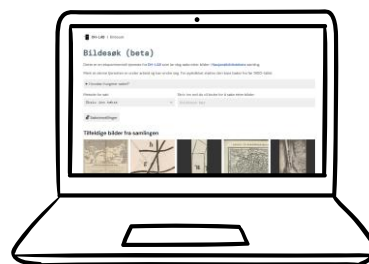




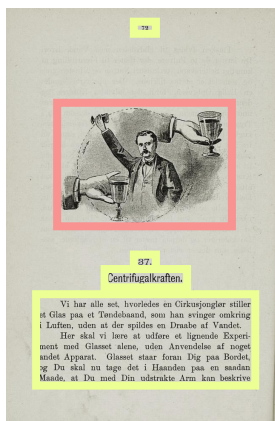
Datagrunnlaget



Bildelikhhet



Søkeapplikasjonen og
veien videre



Datagrunnlaget



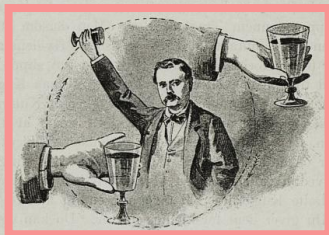
Bildelikhhet



Søkeapplikasjonen og
veien videre

Som del av digitaliseringsløypa kjøres automatisk strukturanalyse av skannede sider

72



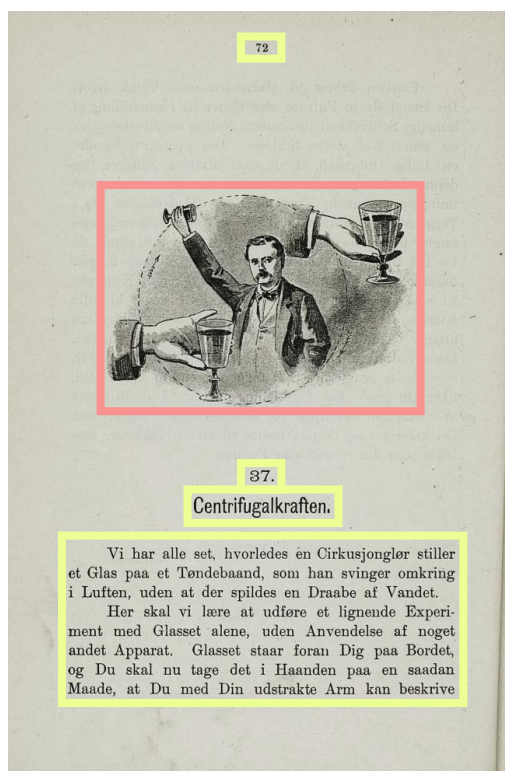
87.

Centrifugalkraften.

Vi har alle set, hvorledes en Cirkusjonglør stiller et Glas paa et Tondebaand, som han svinger omkring i Luften, uden at der spildes en Draabe af Vandet.

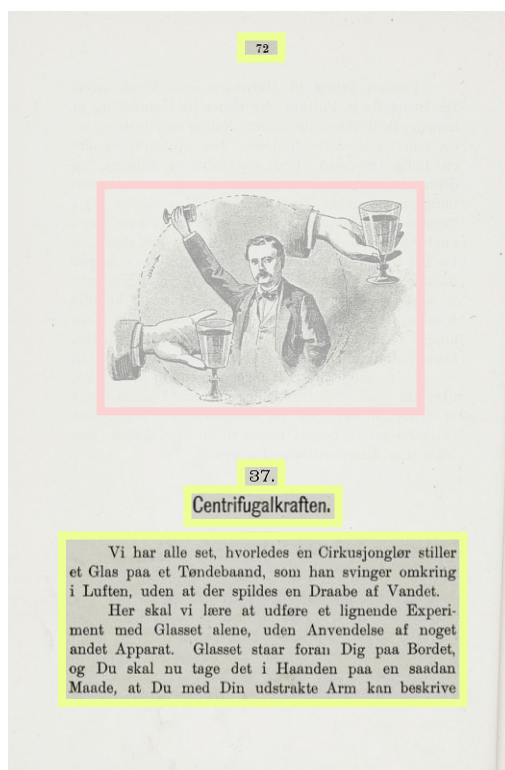
Her skal vi lære at udføre et lignende Experiment med Glasset alene, uden Anvendelse af noget andet Apparat. Glasset staar foran Dig paa Bordet, og Du skal nu tage det i Haanden paa en saadan Maade, at Du med Din udstrakte Arm kan beskrive

Som del av digitaliseringsløypa kjøres automatisk strukturanalyse av skannede sider



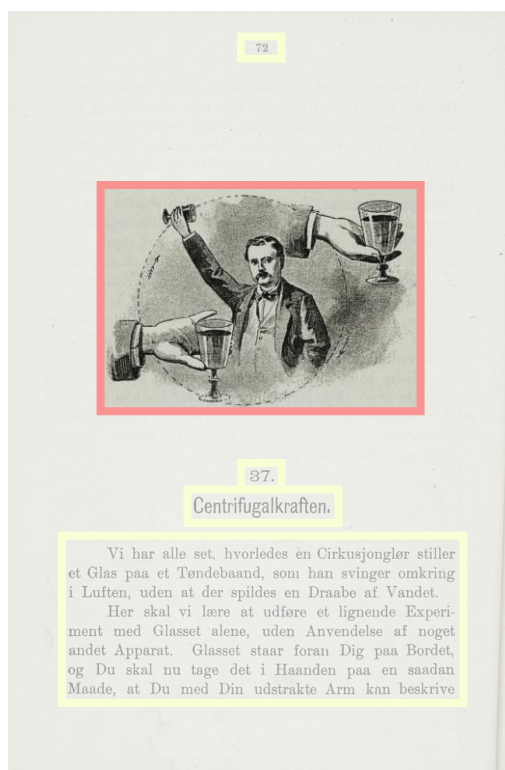
```
[...]  
<Layout>  
  <Page ID="P230" PHYSICAL_IMG_NR="230" HEIGHT="1790" WIDTH="1172" [...] >  
    [...]   
    <PrintSpace ID="P262_PS00001" [...] >  
      [...]   
      <TextBlock ID="P230_TB00003" [...] >  
        <TextLine ID="P230_TL00003" [...] >  
          <String ID="P230_ST00003" CONTENT="Centrifugal" [...] />  
          <SP ID="P230_SP00001" [...] />  
          <String ID="P230_ST00004" CONTENT="kraften," [...] />  
        </TextLine>  
      </TextBlock>  
      [...]   
      <ComposedBlock ID="P230_CB00001" [...] TYPE="Illustration">  
        <GraphicalElement ID="P230_CB00001_SUB" [...] />  
      </ComposedBlock>  
    </PrintSpace>  
  </Page>  
</Layout>
```


Som del av digitaliseringsløypa kjøres automatisk strukturanalyse av skannede sider



```
[...]  
<Layout>  
  <Page ID="P230" PHYSICAL_IMG_NR="230" HEIGHT="1790" WIDTH="1172" [...] >  
    [...]   
    <PrintSpace ID="P262_PS00001" [...] >  
      [...]   
      <TextBlock ID="P230_TB00003" [...] >  
        <TextLine ID="P230_TL00003" [...] >  
          <String ID="P230_ST00003" CONTENT="Centrifugal" [...] />  
          <SP ID="P230_SP00001" [...] />  
          <String ID="P230_ST00004" CONTENT="kraften," [...] />  
        </TextLine>  
      </TextBlock>  
      [...]   
      <ComposedBlock ID="P230_CB00001" [...] TYPE="Illustration">  
        <GraphicalElement ID="P230_CB00001_SUB" [...] />  
      </ComposedBlock>  
    </PrintSpace>  
  </Page>  
</Layout>
```

Som del av digitaliseringsløypa kjøres automatisk strukturanalyse av skannede sider



```
[...]  
<Layout>  
  <Page ID="P230" PHYSICAL_IMG_NR="230" HEIGHT="1790" WIDTH="1172" [...] >  
    [...]   
    <PrintSpace ID="P262_PS00001" [...] >  
      [...]   
      <TextBlock ID="P230_TB00003" [...] >  
        <TextLine ID="P230_TL00003" [...] >  
          <String ID="P230_ST00003" CONTENT="Centrifugal" [...] />  
          <SP ID="P230_SP00001" [...] />  
          <String ID="P230_ST00004" CONTENT="kraften," [...] />  
        </TextLine>  
      </TextBlock>  
      [...]   
      <ComposedBlock ID="P230_CB00001" [...] TYPE="Illustration">  
        <GraphicalElement ID="P230_CB00001_SUB" [...] />  
      </ComposedBlock>  
    </PrintSpace>  
  </Page>  
</Layout>
```

Vi kan bruke IIIF endepunktet til å hente bilder basert på ALTO-XML-koordinatene

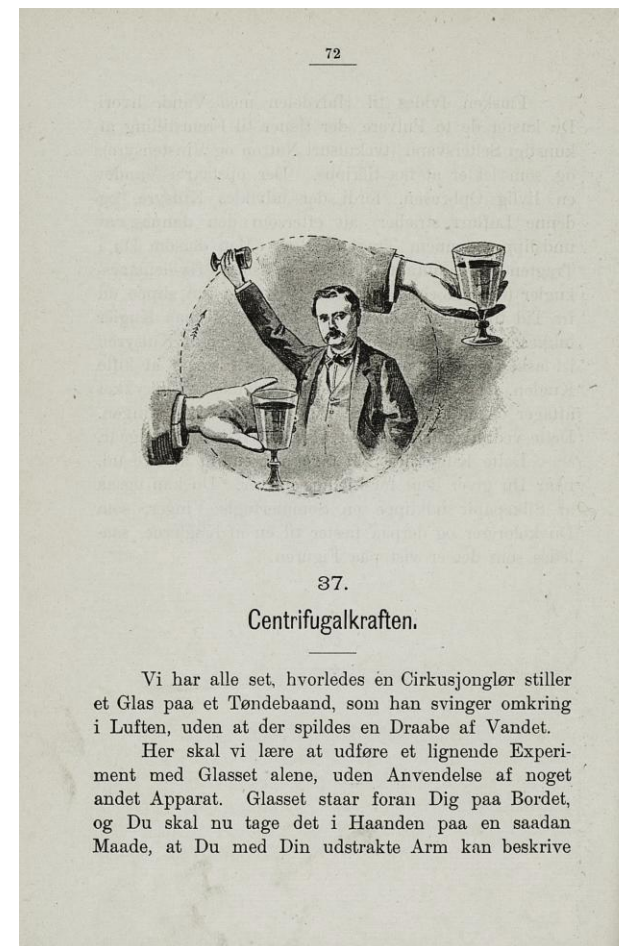
IIIF format:

```
https://nb.no/services/image/resolver/  
  <URN>/  
  <x>,<y>,<w>,<h>/  
  full/0/default.jpg
```

Vi kan bruke IIIF endepunktet til å hente bilder basert på ALTO-XML-koordinatene

IIIF format:

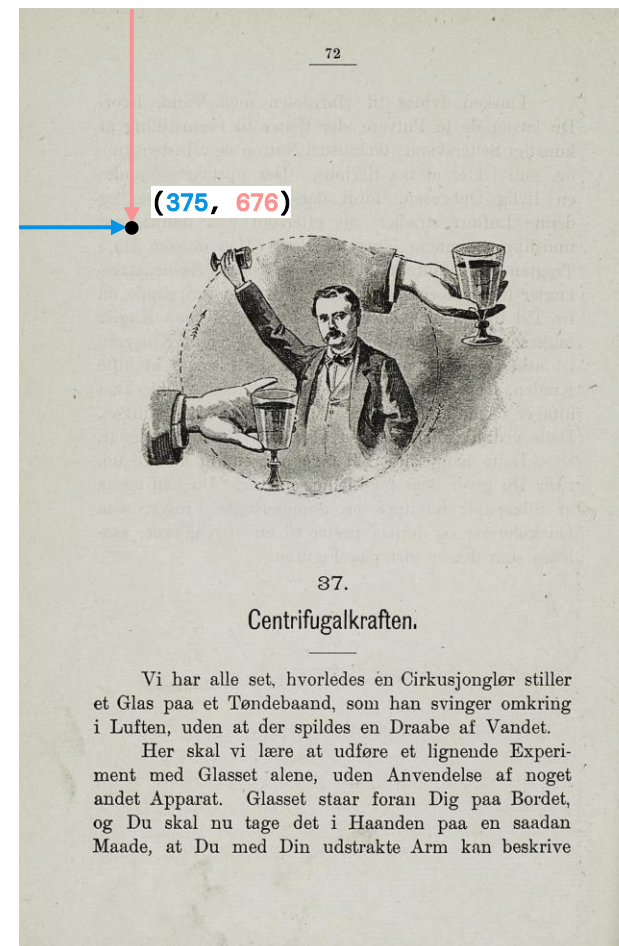
[https://nb.no/services/image/resolver/
URN:NBN:no-nb_digibok_2009041503013_0230/
<x>,<y>,<w>,<h>/
full/0/default.jpg](https://nb.no/services/image/resolver/URN:NBN:no-nb_digibok_2009041503013_0230/<x>,<y>,<w>,<h>/full/0/default.jpg)



Vi kan bruke IIIF endepunktet til å hente bilder basert på ALTO-XML-koordinatene

IIIF format:

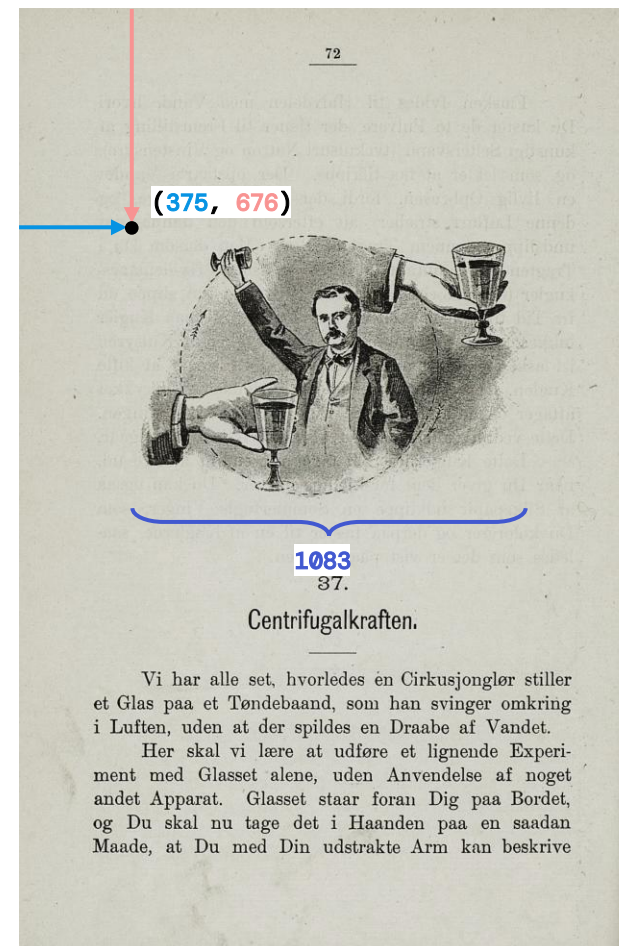
[https://nb.no/services/image/resolver/
URN:NBN:no-nb_digibok_2009041503013_0230/
375,676,<w>,<h>/
full/0/default.jpg](https://nb.no/services/image/resolver/URN:NBN:no-nb_digibok_2009041503013_0230/375,676,<w>,<h>/full/0/default.jpg)



Vi kan bruke IIIF endepunktet til å hente bilder basert på ALTO-XML-koordinatene

IIIF format:

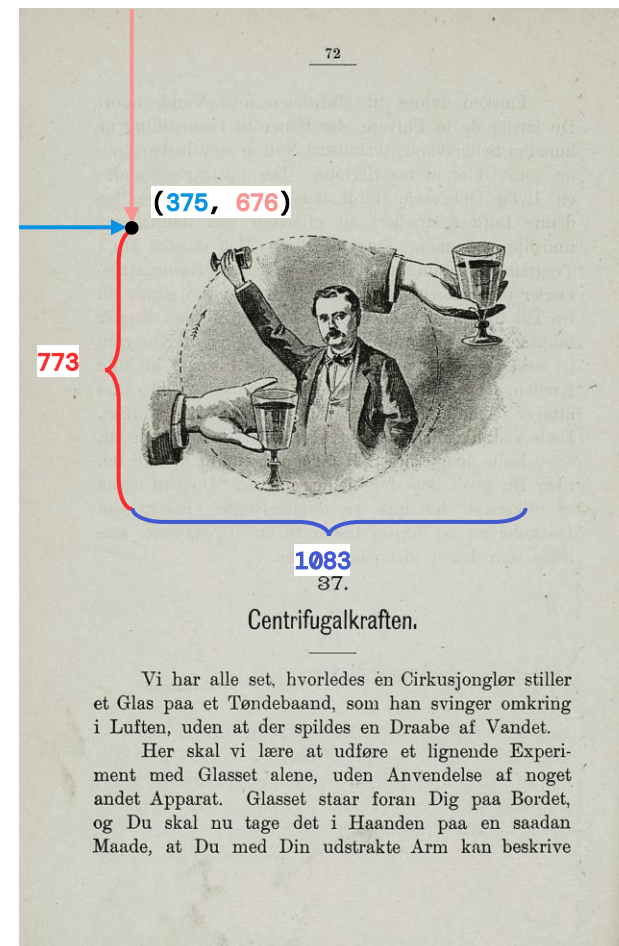
[https://nb.no/services/image/resolver/
URN:NBN:no-nb_digibok_2009041503013_0230/
375,676,1083,<h>/
full/0/default.jpg](https://nb.no/services/image/resolver/URN:NBN:no-nb_digibok_2009041503013_0230/375,676,1083,<h>/full/0/default.jpg)



Vi kan bruke IIIF endepunktet til å hente bilder basert på ALTO-XML-koordinatene

IIIF format:

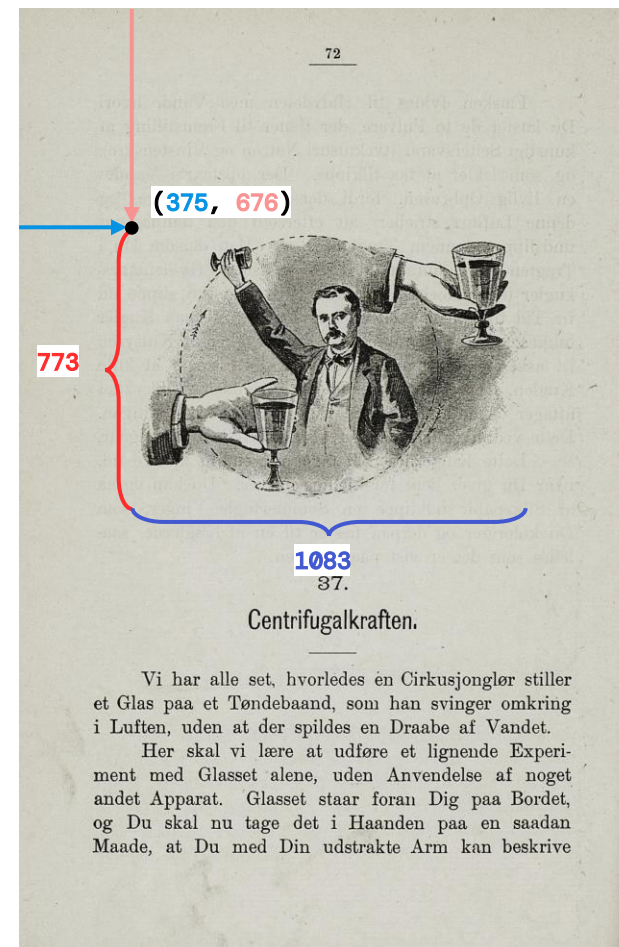
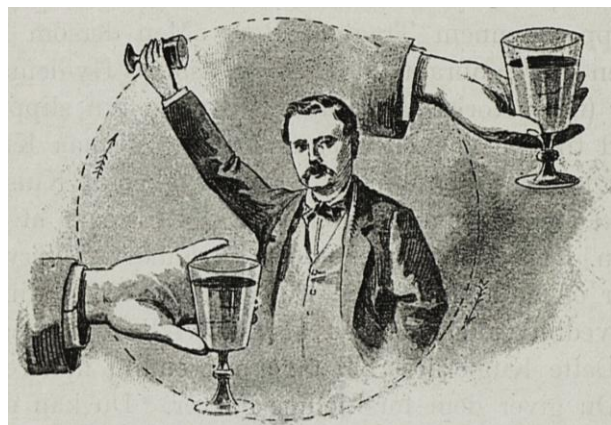
[https://nb.no/services/image/resolver/
URN:NBN:no-nb_digibok_2009041503013_0230/
375,676,1083,773/
full/0/default.jpg](https://nb.no/services/image/resolver/URN:NBN:no-nb_digibok_2009041503013_0230/375,676,1083,773/full/0/default.jpg)

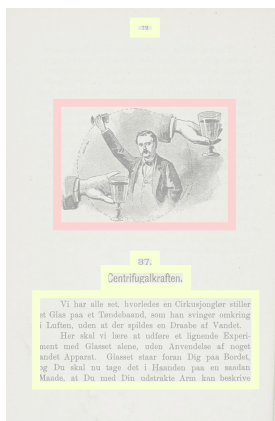


Vi kan bruke IIIF endepunktet til å hente bilder basert på ALTO-XML-koordinatene

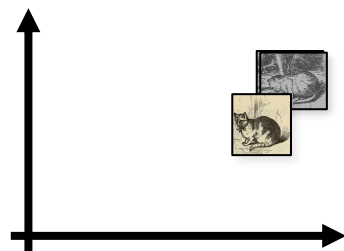
IIIF format:

[https://nb.no/services/image/resolver/
URN:NBN:no-nb_digibok_2009041503013_0230/
375,676,1083,773/
full/0/default.jpg](https://nb.no/services/image/resolver/URN:NBN:no-nb_digibok_2009041503013_0230/375,676,1083,773/full/0/default.jpg)





Datagrunnlaget

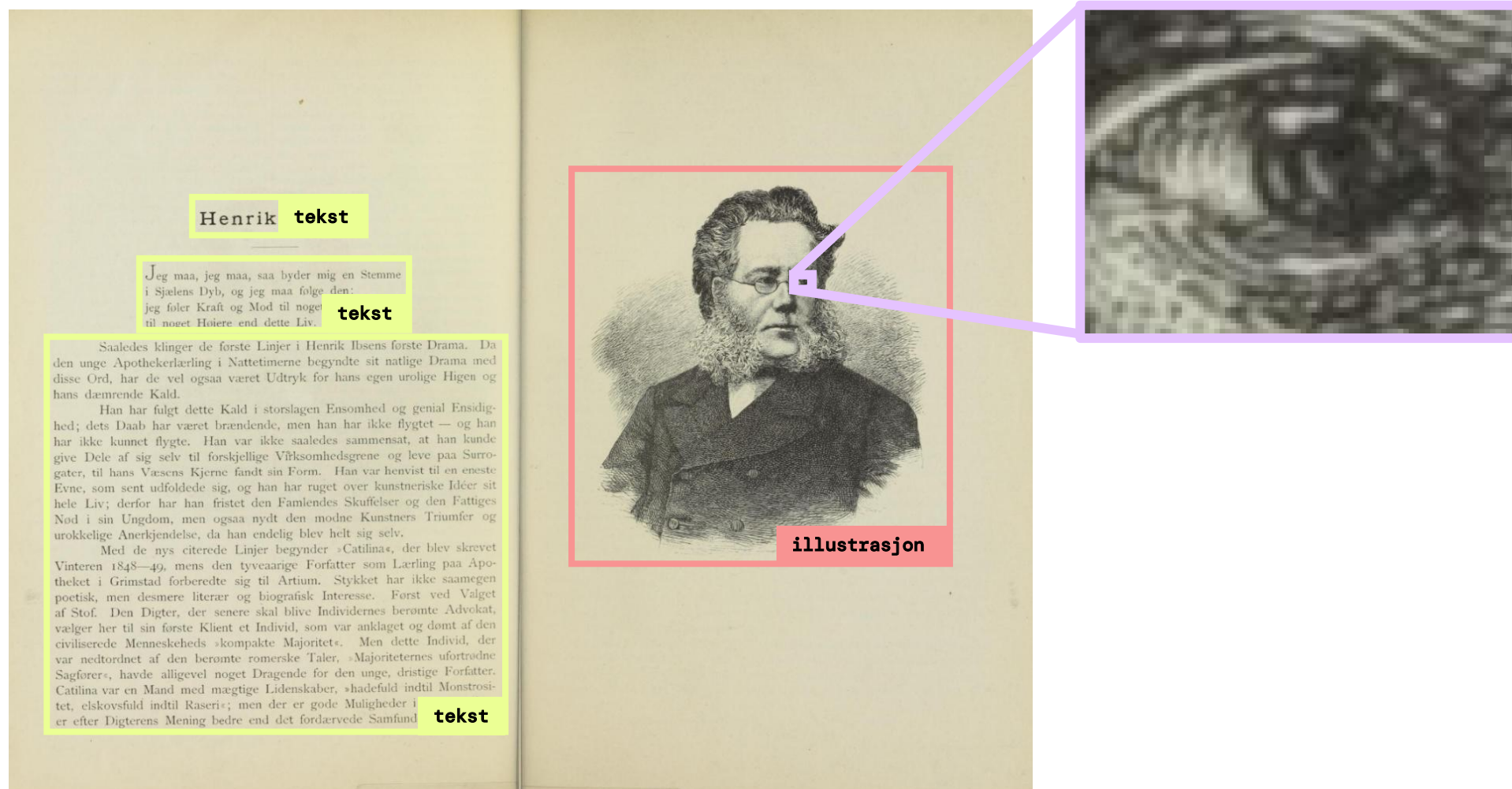


Bildelikhhet



Søkeapplikasjonen og
veien videre

Et digitalt bilde er en tabell med tall



Et digitalt bilde er en tabell med tall

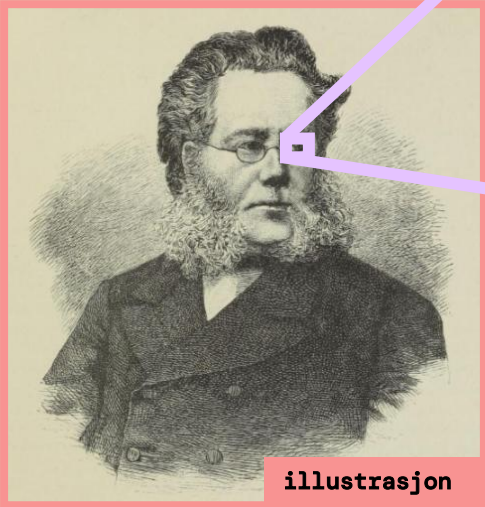
Henrik tekst

Jeg maa, jeg maa, saa byder mig en Stemme
i Sjælens Dyb, og jeg maa følge den:
jeg føler Kraft og Mod til noget
til noget Hoiere end dette Liv. tekst


Saaledes klinger de første Linjer i Henrik Ibsens første Drama. Da den unge Apothekerlærling i Nattetimerne begyndte sit natlige Drama med disse Ord, har de vel ogsaa været Udtryk for hans egen urolige Higen og hans dæmrende Kald.

Han har fulgt dette Kald i storslagen Ensomhed og genial Ensidighed; dets Daab har været brændende, men han har ikke flygtet — og han har ikke kunnet flygte. Han var ikke saaledes sammensat, at han kunde give Dele af sig selv til forskjellige Virksomhedsgrene og leve paa Surrogater, til hans Væscens Kjerne fandt sin Form. Han var henvist til en eneste Evne, som sent udfoldede sig, og han har ruget over kunstneriske Idéer sit hele Liv; derfor har han fristet den Familiendes Skuffelser og den Fattiges Nød i sin Ungdom, men ogsaa nydt den modne Kunstners Triumfer og urokkelige Anerkjendelse, da han endelig blev helt sig selv.

Med de nys citerede Linjer begynder »Catilina«, der blev skrevet Vinteren 1848—49, mens den tyveaarige Forfatter som Lærling paa Apotheket i Grimstad forberedte sig til Artium. Stykket har ikke saamegen poetisk, men desmere literær og biografisk Interesse. Først ved Valget af Stof. Den Digter, der senere skal blive Individernes berømte Advokat, vælger her til sin første Klient et Individ, som var anklaget og dømt af den civiliserede Menneskeheds »kompakte Majoritet«. Men dette Individ, der var nedtordnet af den berømte romerske Taler, »Majoriteternes ufortrødne Sagfører«, havde alligevel noget Dragende for den unge, dristige Forfatter. Catilina var en Mand med mægtige Lidenskaber, »hadefuld indtil Monstrositet, elskovsfuld indtil Raseri«; men der er gode Muligheder i er efter Digterens Mening bedre end det forværrede Samfund tekst

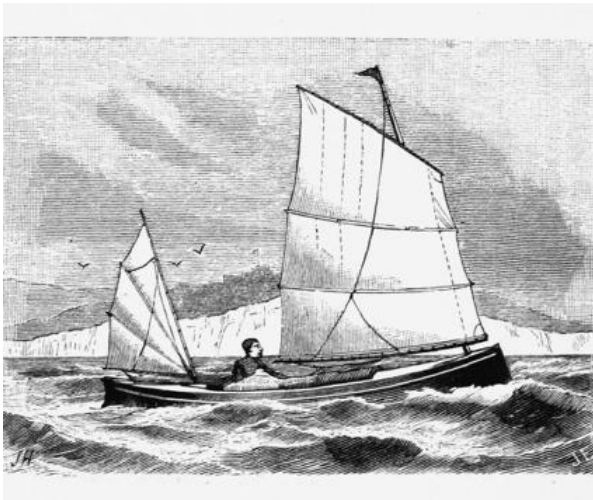


illustrasjon



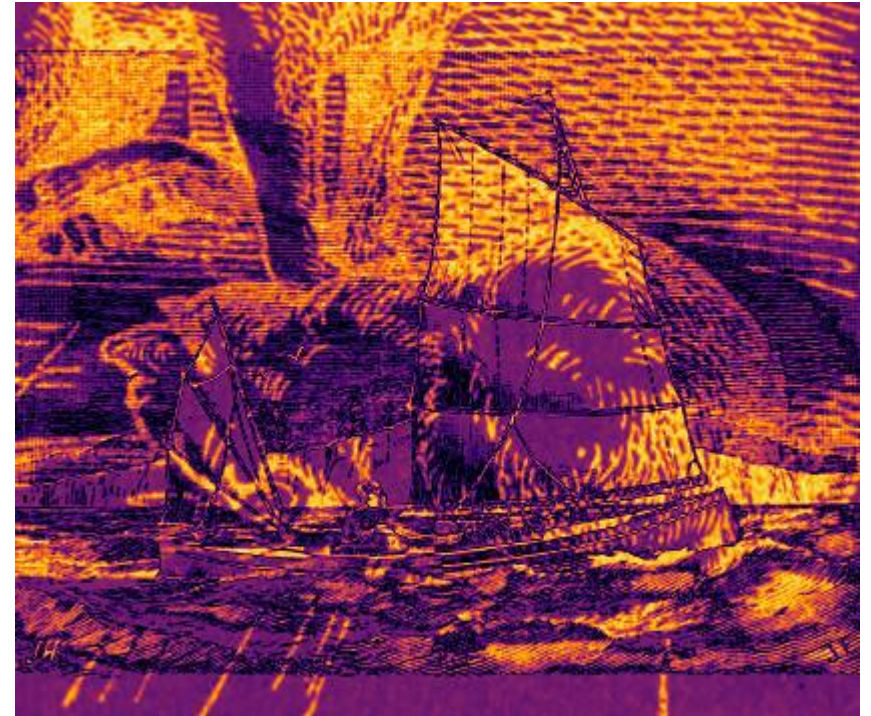
71	62	57	48	54	54	54	66	78	71
112	82	79	78	73	55	54	81	109	88
86	71	71	71	67	56	77	131	136	91
67	62	72	63	51	48	67	79	86	80
147	139	142	90	42	45	48	52	57	59
116	97	88	62	54	70	64	70	64	58
58	52	52	49	50	60	77	74	59	70
57	57	48	45	43	49	72	103	75	63
97	81	51	44	50	56	64	100	83	59
103	101	65	63	59	57	66	69	69	75

Vi kan bruke bildedifferansen for å måle likhet mellom bilder

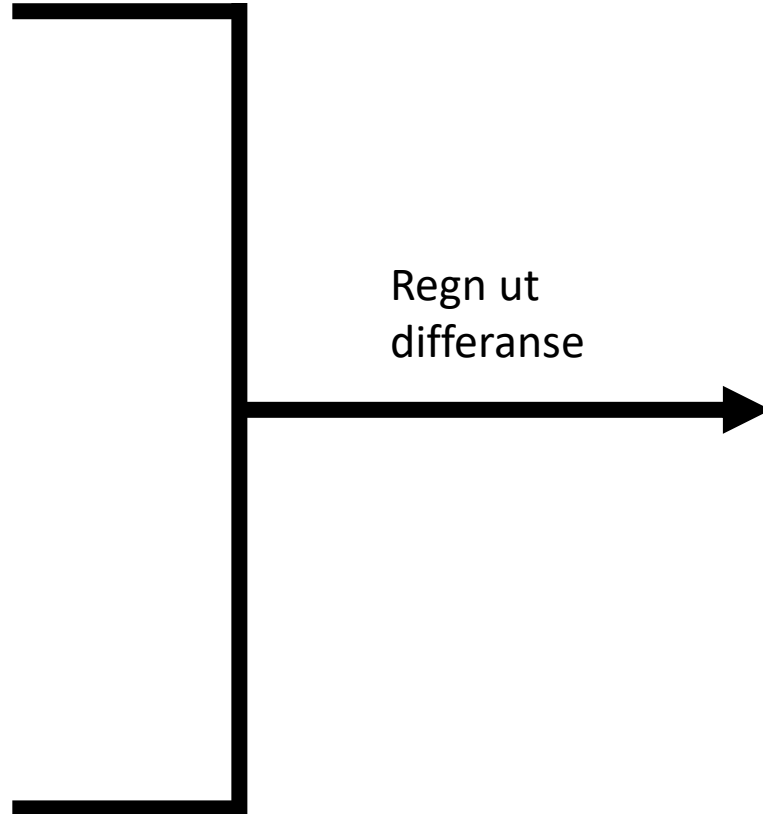
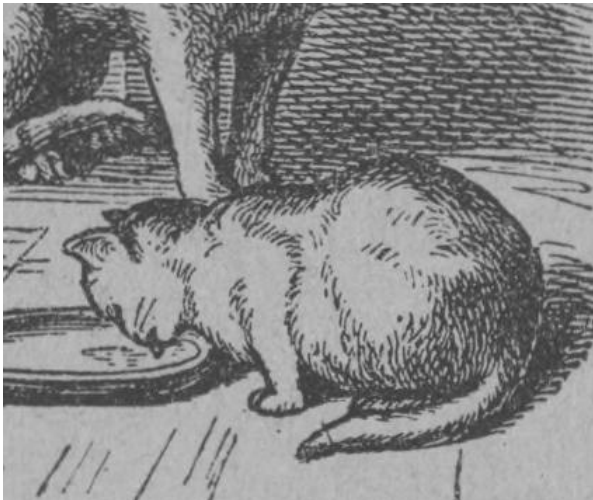


Regn ut
differanse

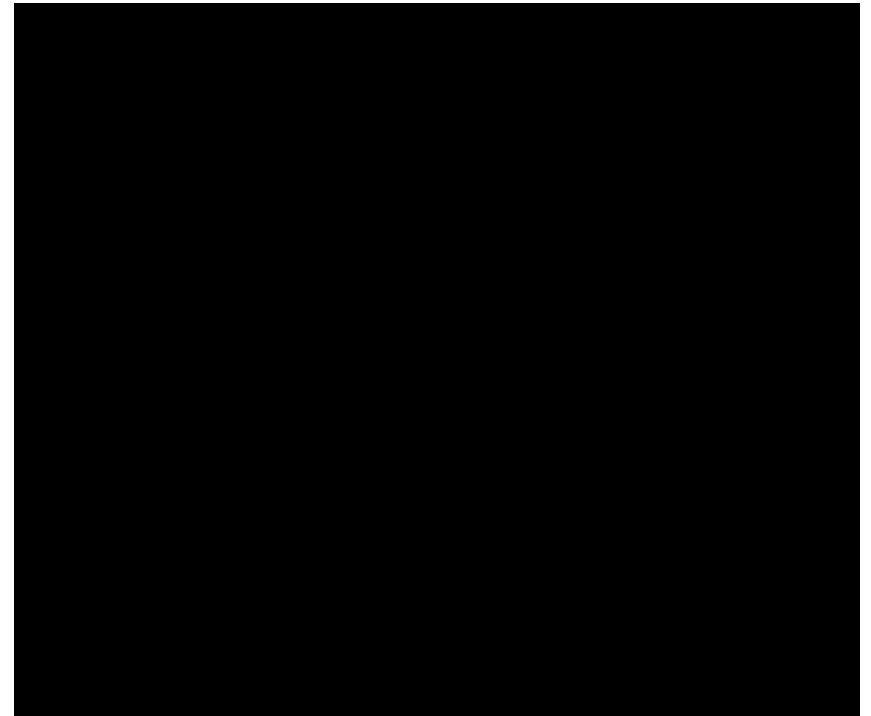
Helt forskjellige!



Vi kan bruke bildedifferansen for å måle
likhet mellom bilder



Like!



Vi kan bruke bildedifferansen for å måle likhet mellom bilder

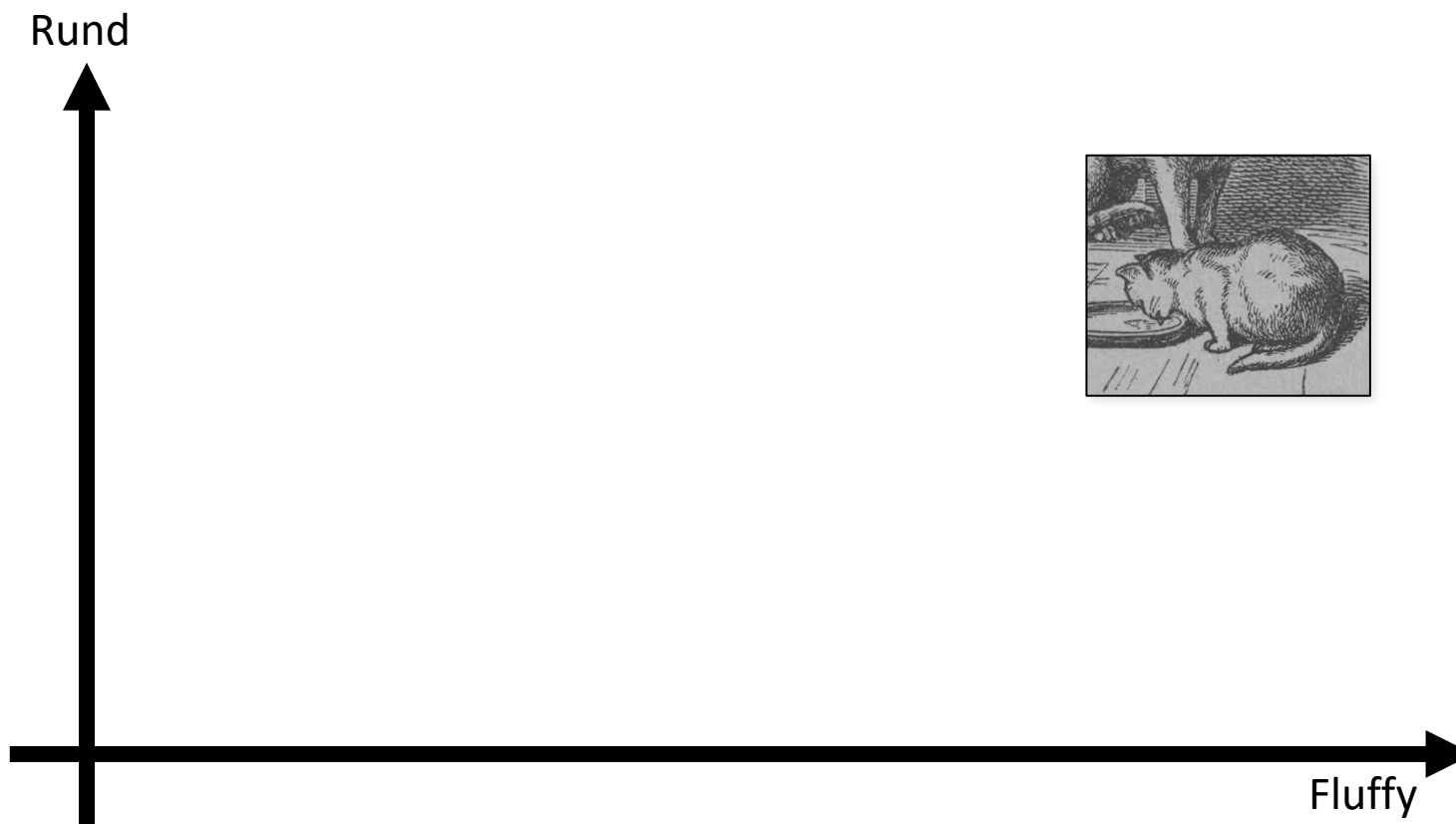


Regn ut
differanse

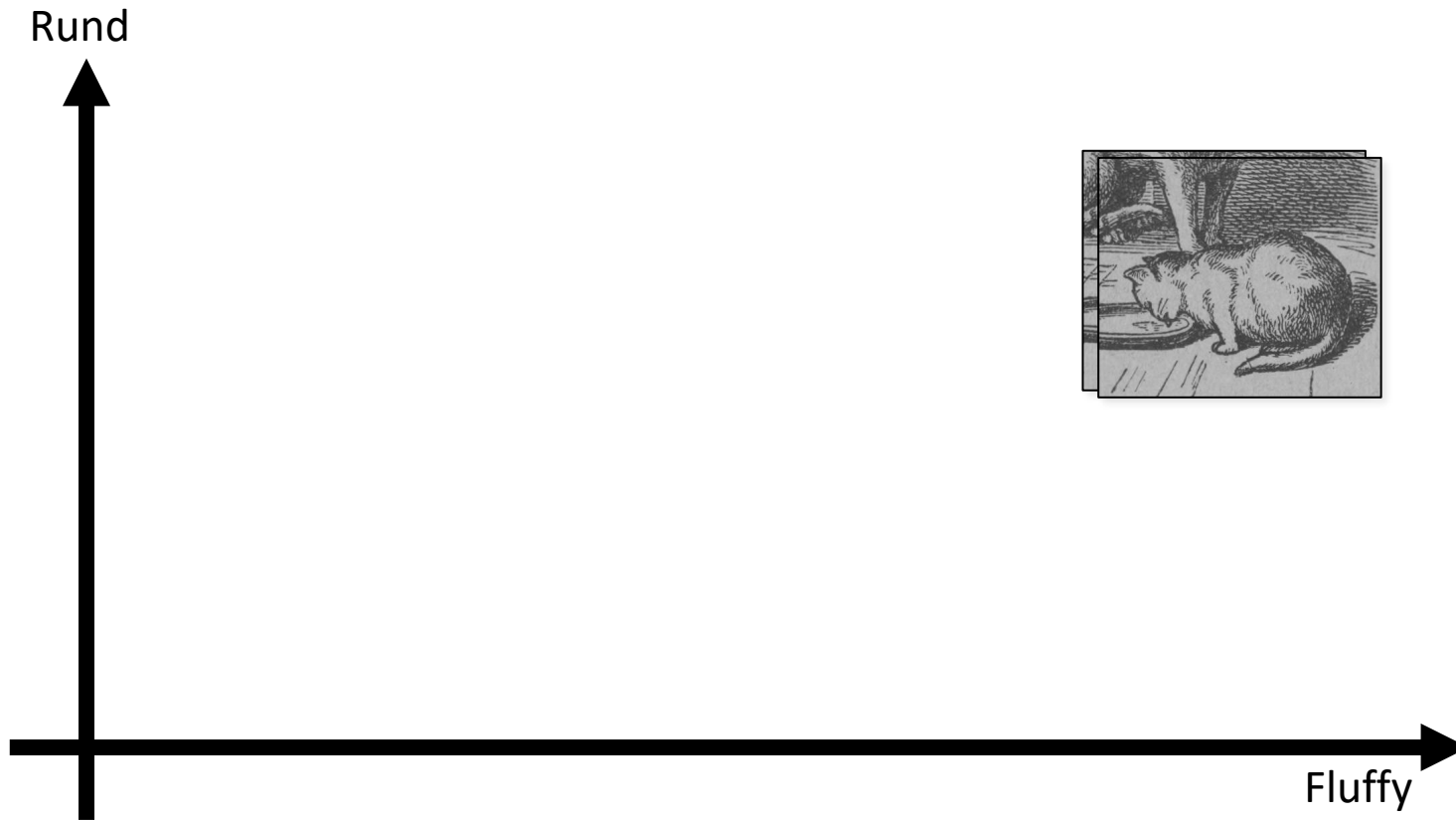
Veldig forskjellige!



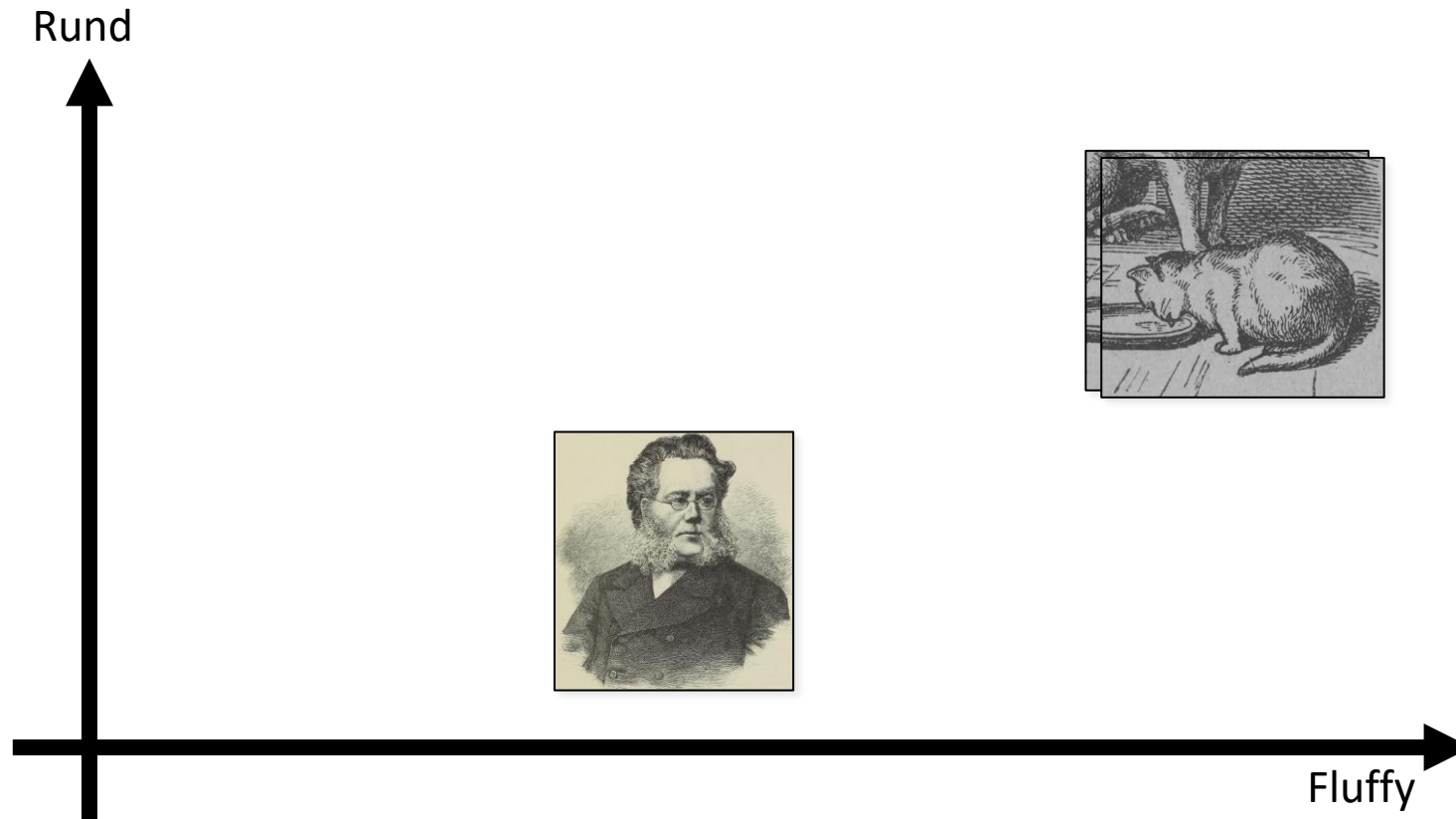
En mer robust metodikk henter ut semantiske
bildeegenskaper som vi kan sammenlikne



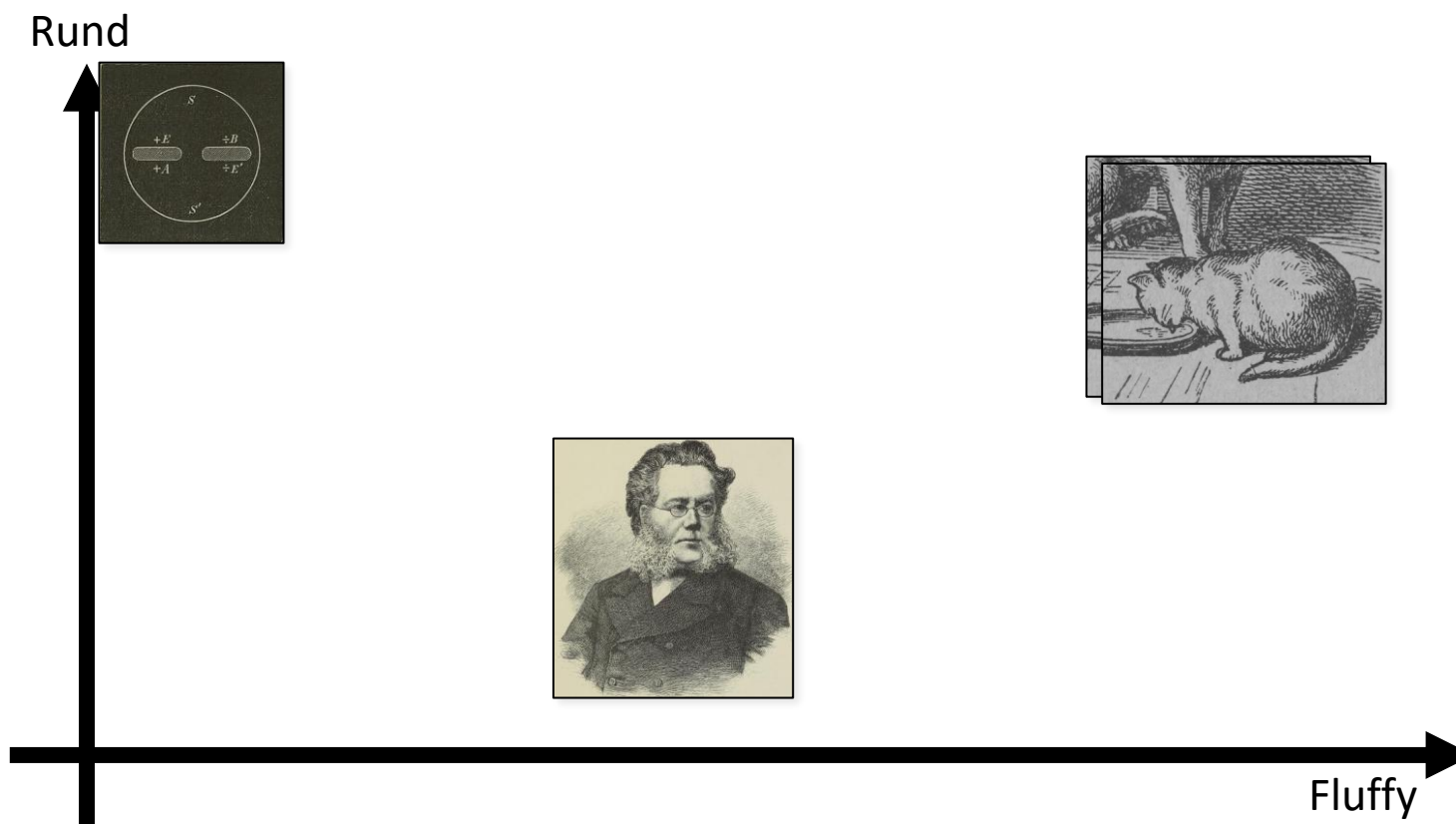
En mer robust metodikk henter ut semantiske
bildeegenskaper som vi kan sammenlikne



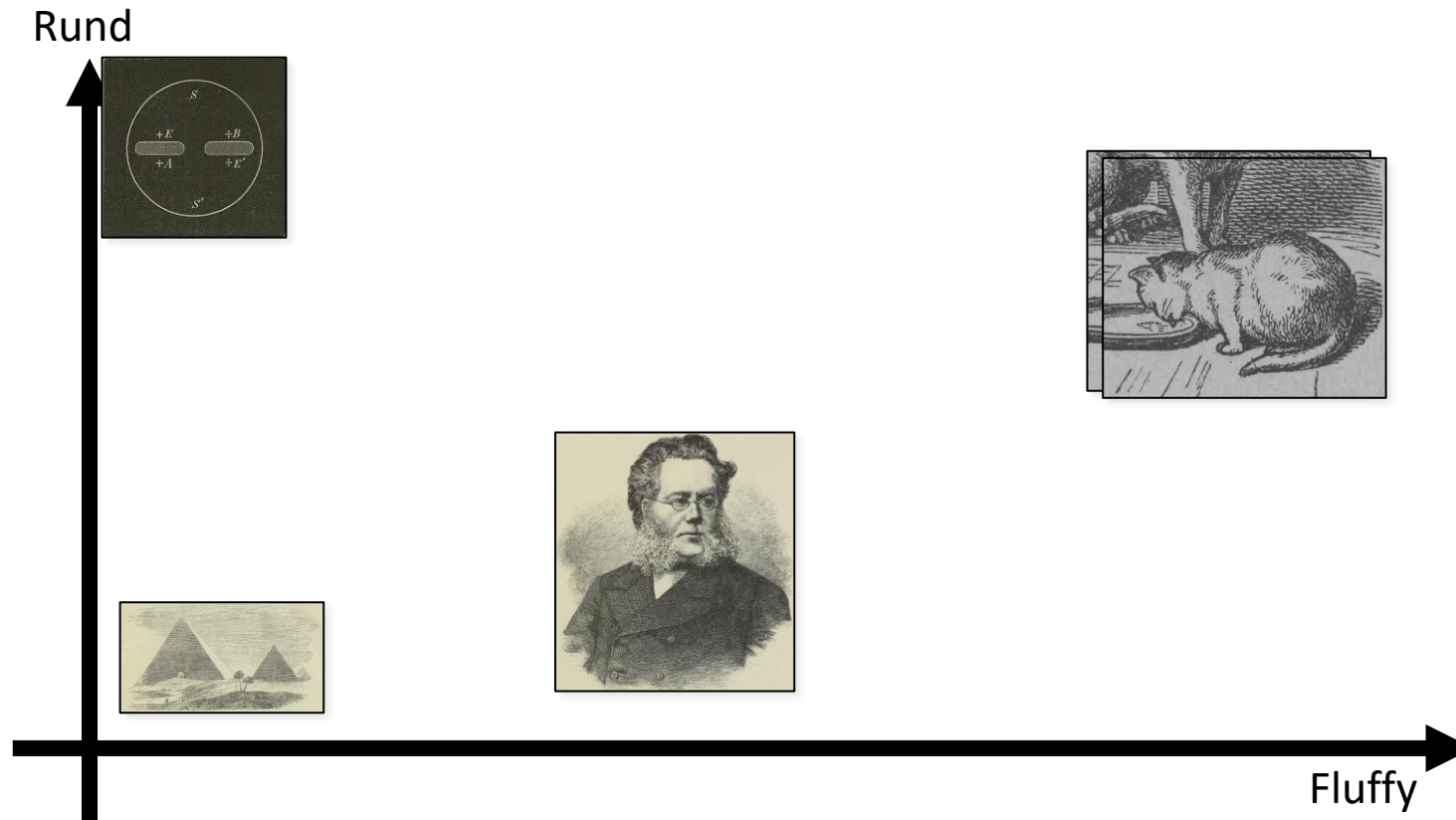
En mer robust metodikk henter ut semantiske bildeegenskaper som vi kan sammenlikne



En mer robust metodikk henter ut semantiske bildeegenskaper som vi kan sammenlikne



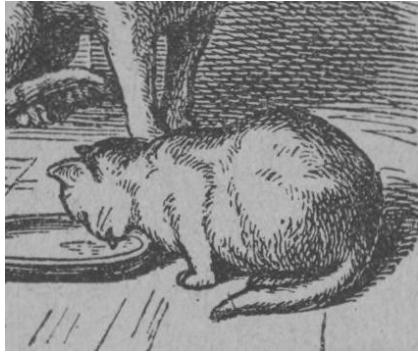
En mer robust metodikk henter ut semantiske bildeegenskaper som vi kan sammenlikne



En mer robust metodikk henter ut semantiske bildeegenskaper som vi kan sammenlikne



En måte å finne slike egenskaper er ved å bruke ferdigtrente KI-modeller

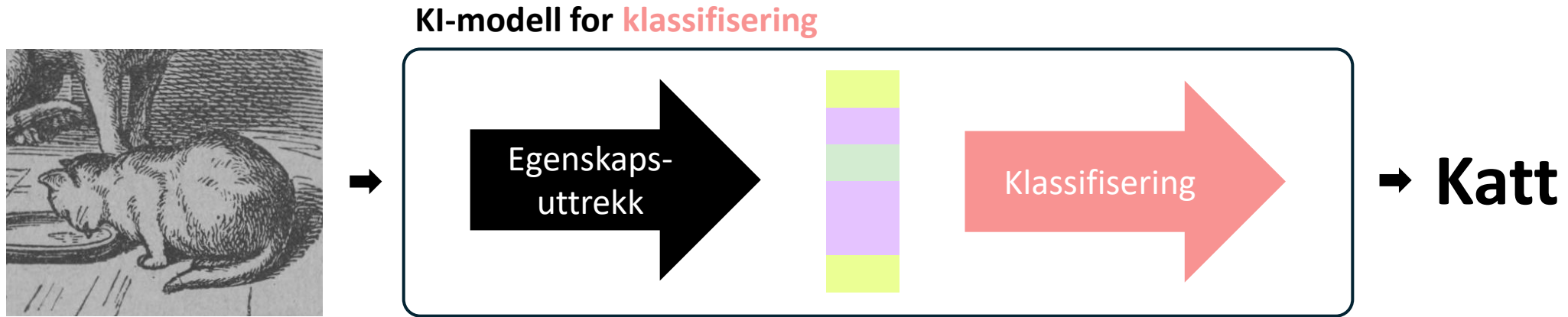


KI-modell for **klassifisering**

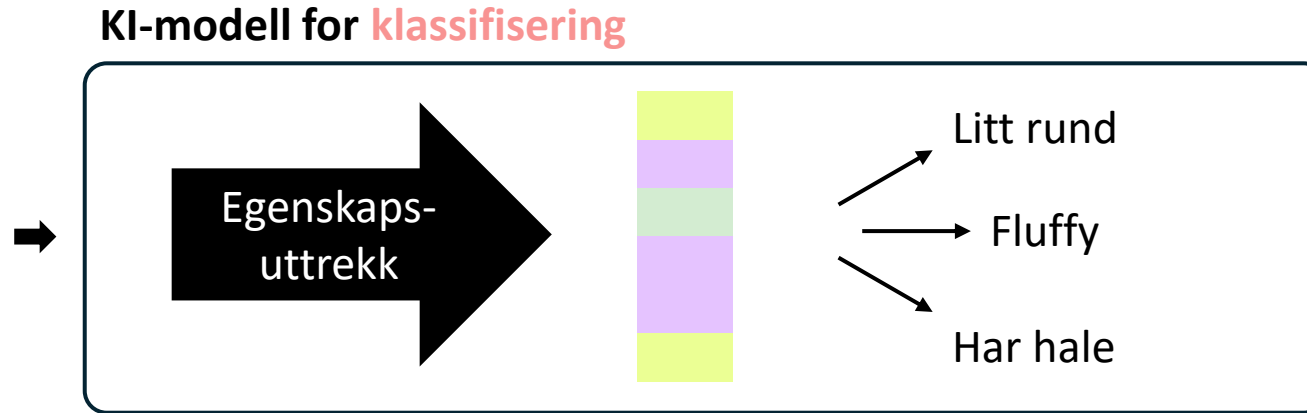
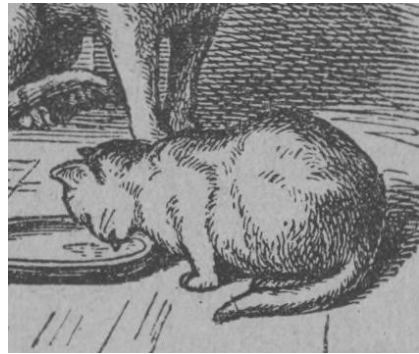


→ **Katt**

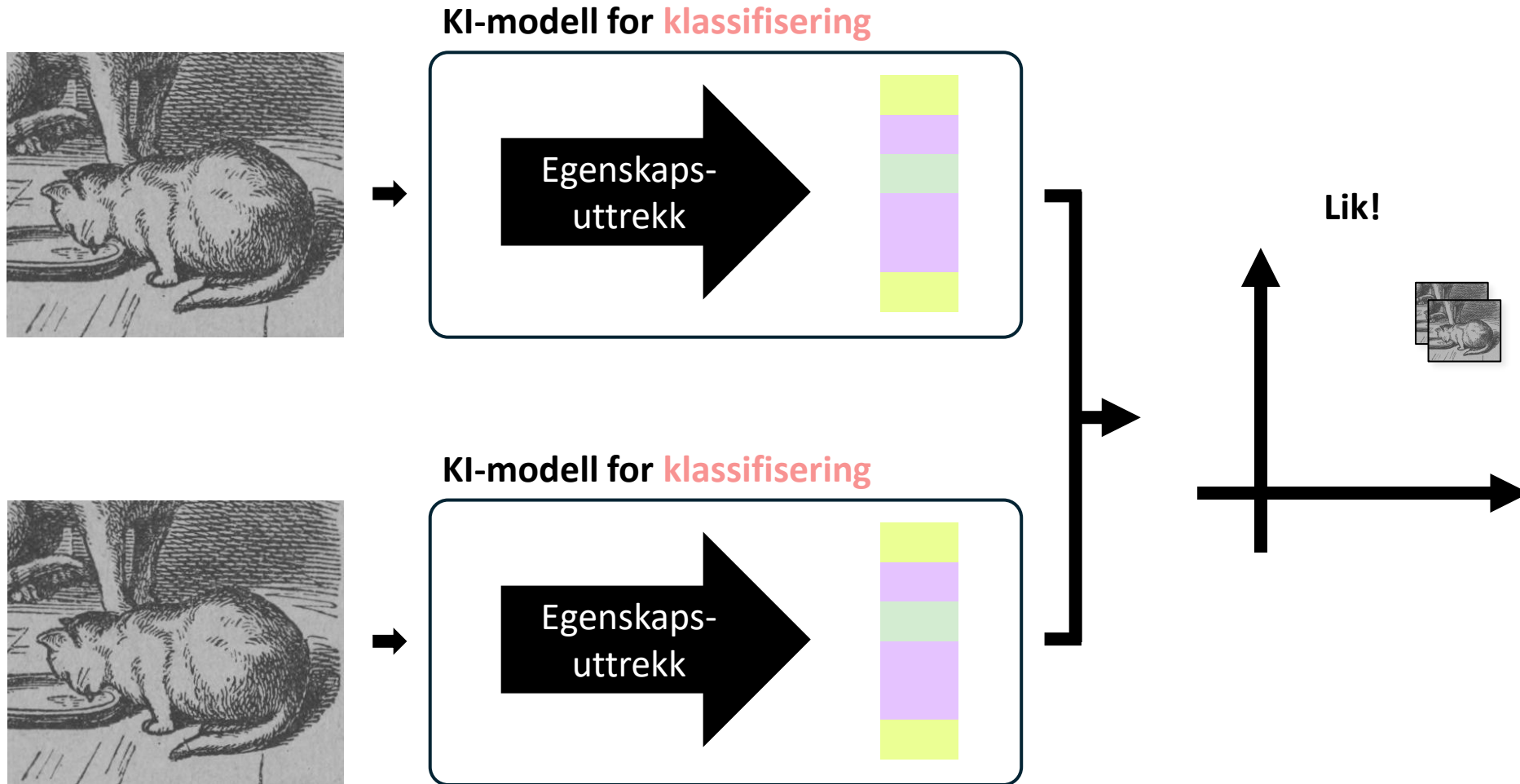
En måte å finne slike egenskaper er ved å bruke ferdigtrente KI-modeller



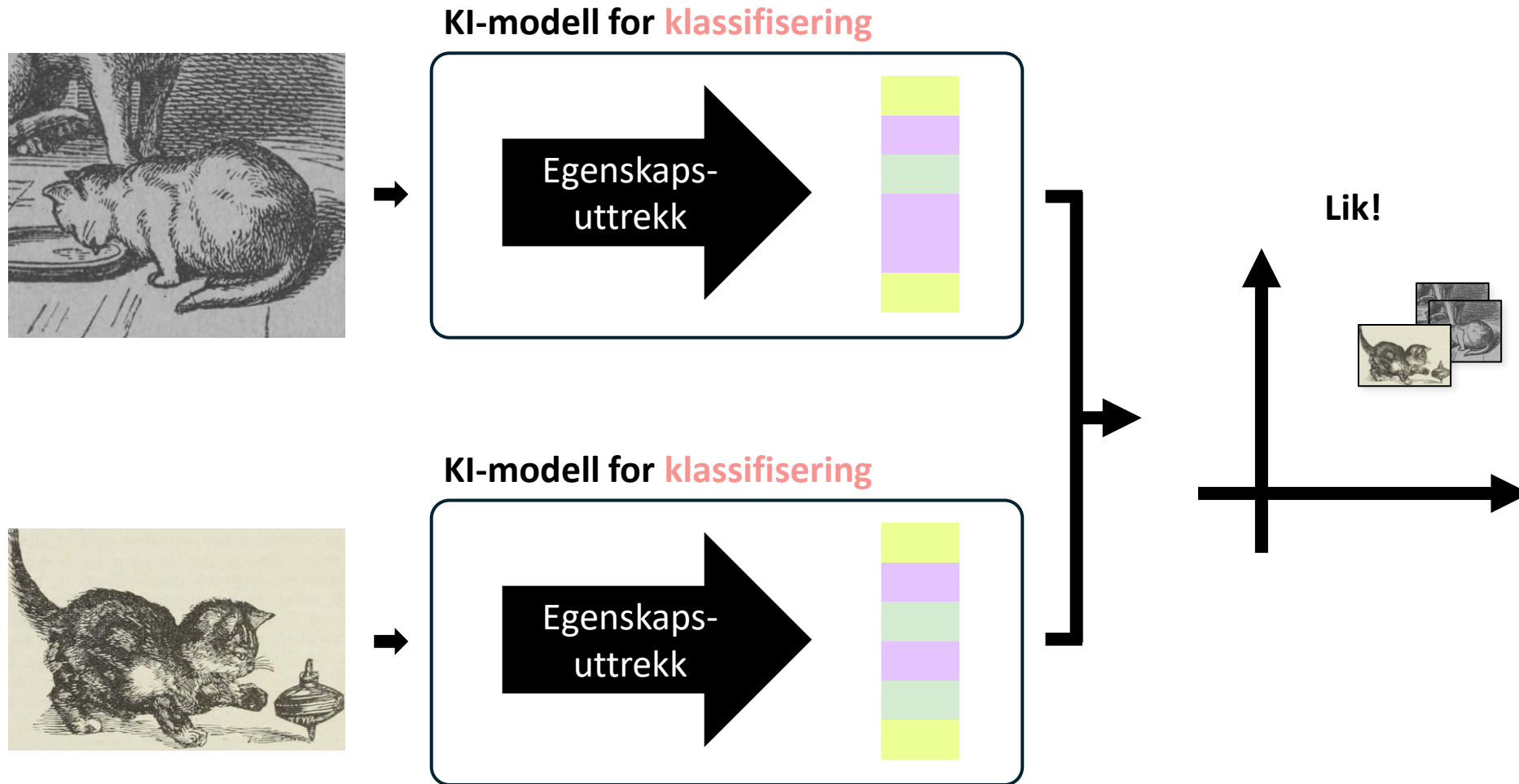
En måte å finne slike egenskaper er ved å bruke ferdigtrente KI-modeller

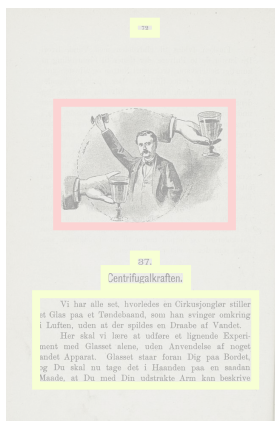


En måte å finne slike egenskaper er ved å bruke ferdigtrente KI-modeller



En måte å finne slike egenskaper er ved å bruke ferdigtrente KI-modeller

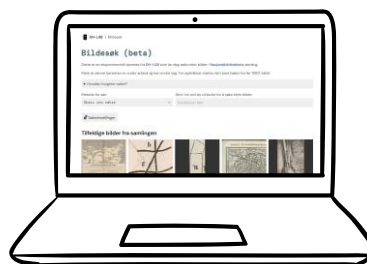




Datagrunnlaget



Bildelikhhet



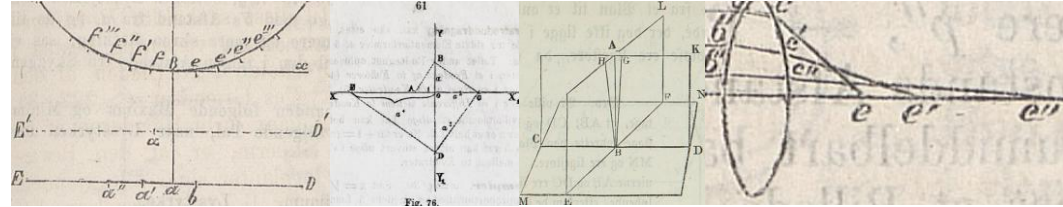
Søkeapplikasjonen og
veien videre

Vi har bygd en prototype bildesøkapplikasjon
med ca 400 000 bilder fra bøker fra før 1900



Bilderepresentasjoner fra KI-modeller kan brukes for å automatisk gruppere bilder

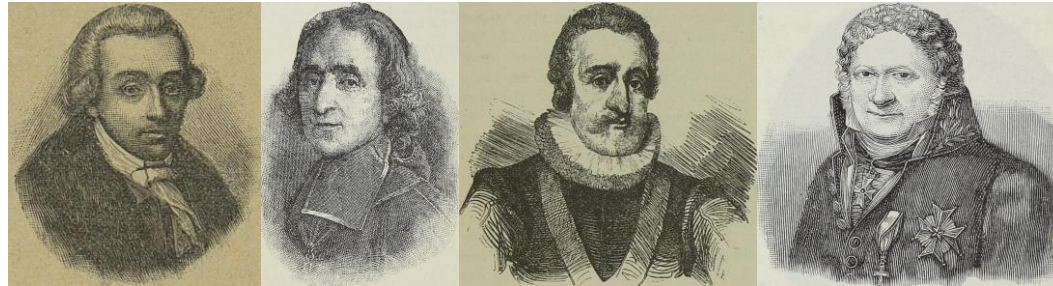
Matematiske figurer



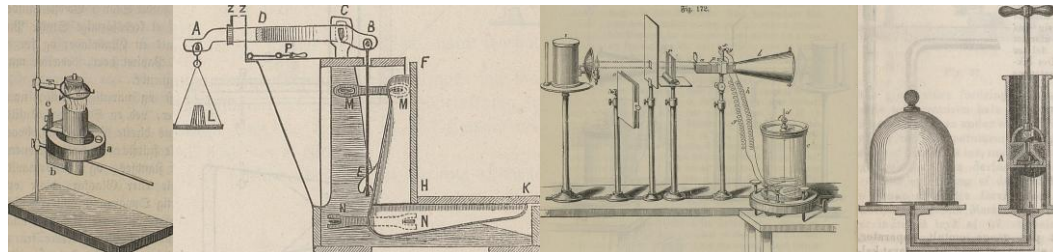
Dekorative element



Portrettillustrasjoner



Ingeniørtegninger



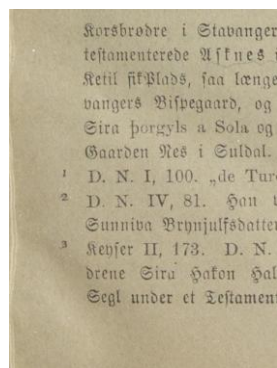
Vi publiserte en artikkel om bildesøkappen på CHR konferansen i fjor



<https://arxiv.org/pdf/2410.14969>

<https://github.com/sprakbanken/CHR24-image-retrieval/>

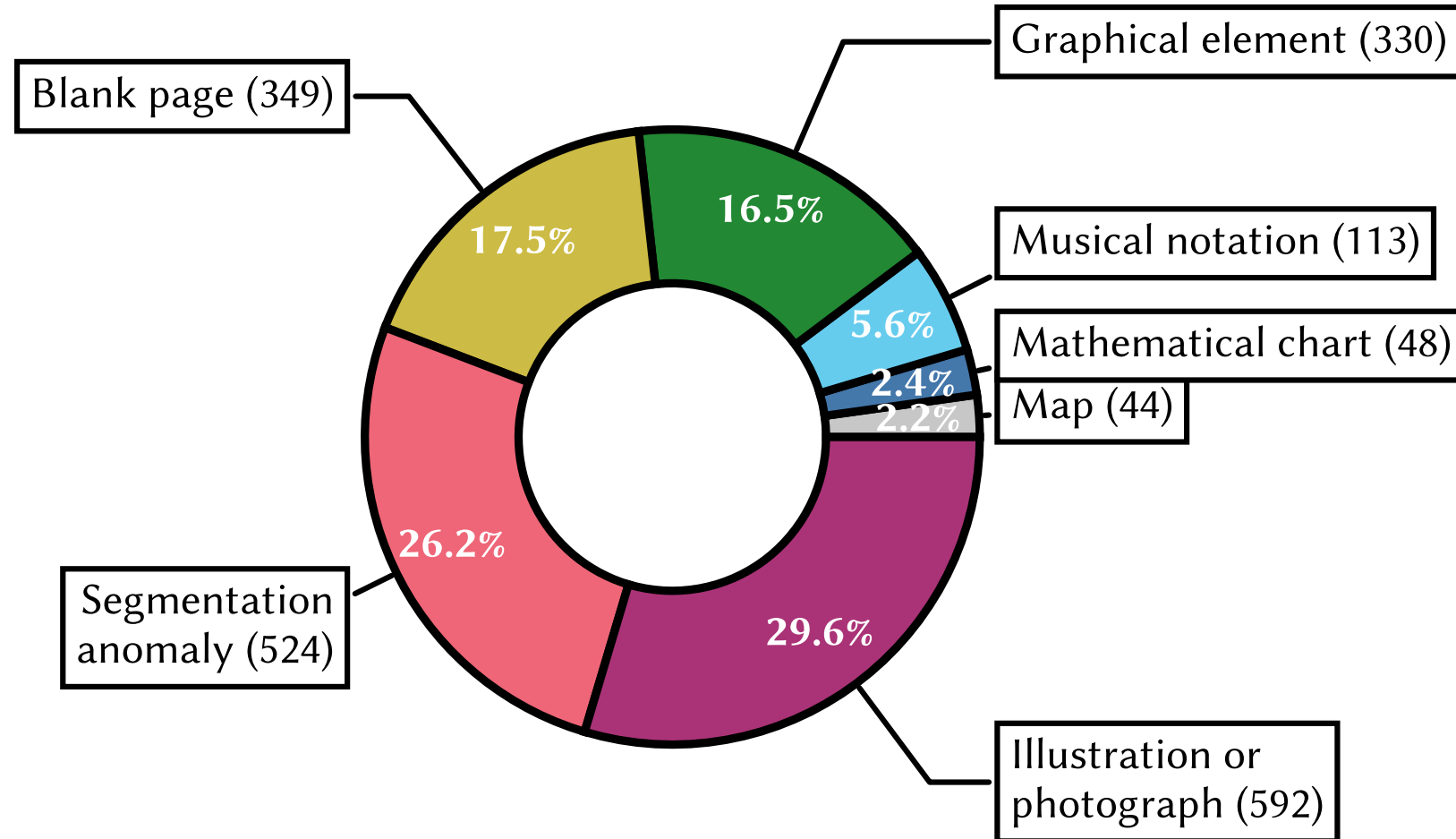
Mange ikke-tekst-elementer er
segmenteringsfeil eller blanke sider



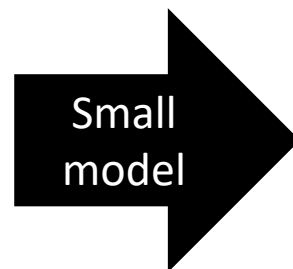
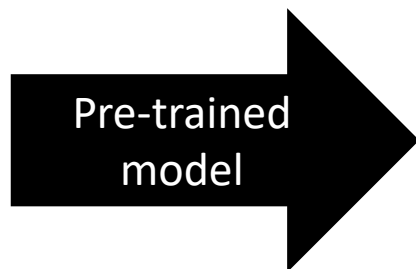
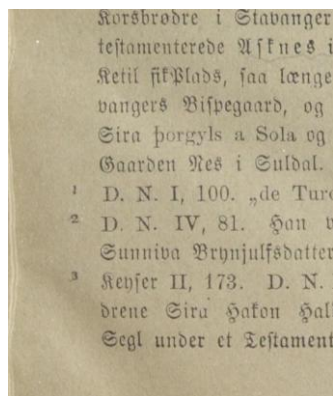
		Dis. Dec. 1822				
1	Godt.	1,419	1	2	17	Sophieslund.
1		1,541	1	3	8	
1		1,247	1	1	18	Solyst.
1		1,504	1	3	4	
1		1,407	1	1	-	
1		53	-	-	7	Aunahoug.
heri Spd. lug.						
320		6,358	6	4	13	
1		1,445	1	2	30	Vennersborg.
en Spd. lug.						
180		2,824	3	-	8	
1		550	-	3	-	
en Spd. lug.	Mindre godt.					
320		5,321	5	3	22	
1		46	-	6		Bynd.
1	Godt.					
500		5,000	5	3	4	Spigervirket.

0	6	
7	30	
8	30	
0	32	
n 91	6	
5	19	
8	24	
1	31	
10	675	1
0	156	
0	11	
id 1	155	
61	997	2
2		
6		
0		

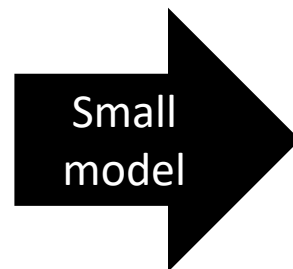
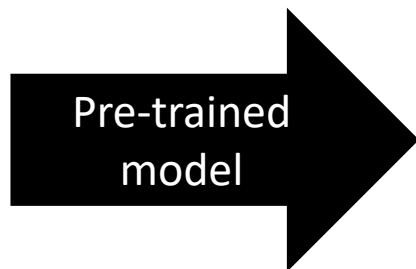
Vi annoterte 2000 bilder



Vi brukte bilderepresentasjonene for å trene en liten KI-modell til å klassifisere bildeelementene



Segmentation error



Illustration

Resultatene viser at vi kan fjerne omtrent 97% av alle blanke sider

True class	Predicted class						
	Segmentation anomaly	Blank page	Graphical element	Illustration or photograph	Musical notation	Map	Mathematical chart
Segmentation anomaly	496	5	28	8	2	1	2
Blank page	11	339	8	1	0	0	0
Graphical element	14	2	278	15	1	0	2
Illustration or photograph	1	3	16	558	1	2	5
Musical notation	1	0	0	0	109	0	0
Map	1	0	0	2	0	41	0
Mathematical chart	0	0	0	8	0	0	39

De fleste* «falske positive» blanke sider inneholder kun stempel

1 Illustration

1 Bar code

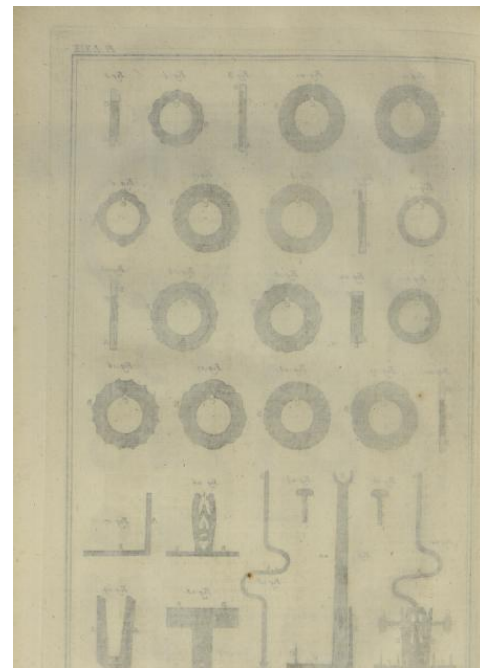
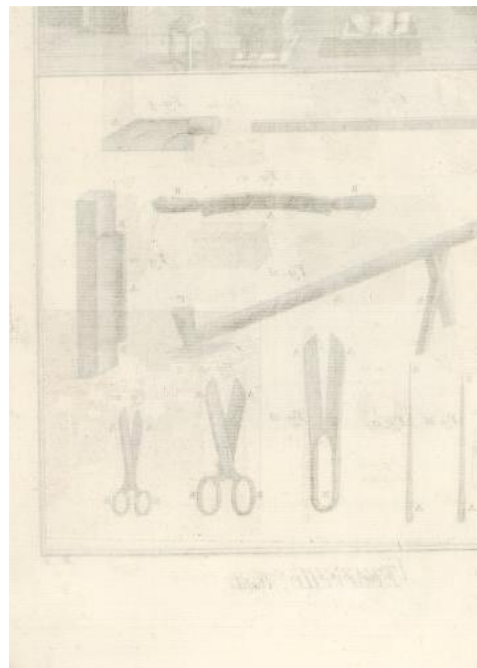
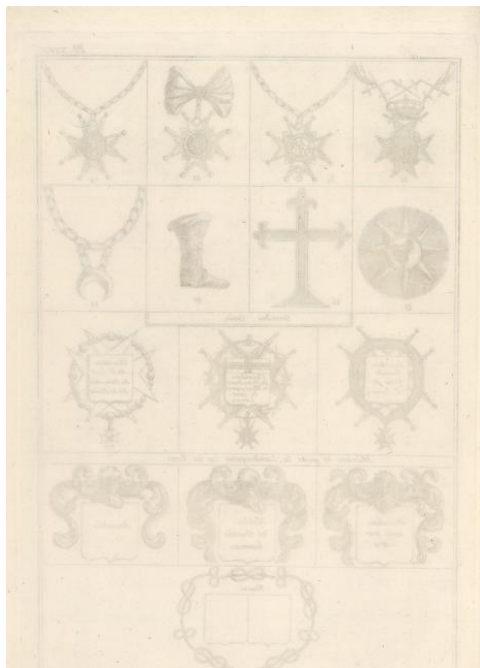
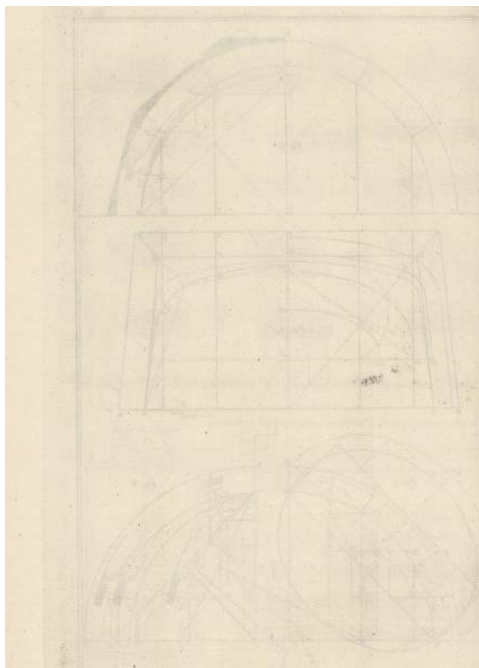


1 Decorative element

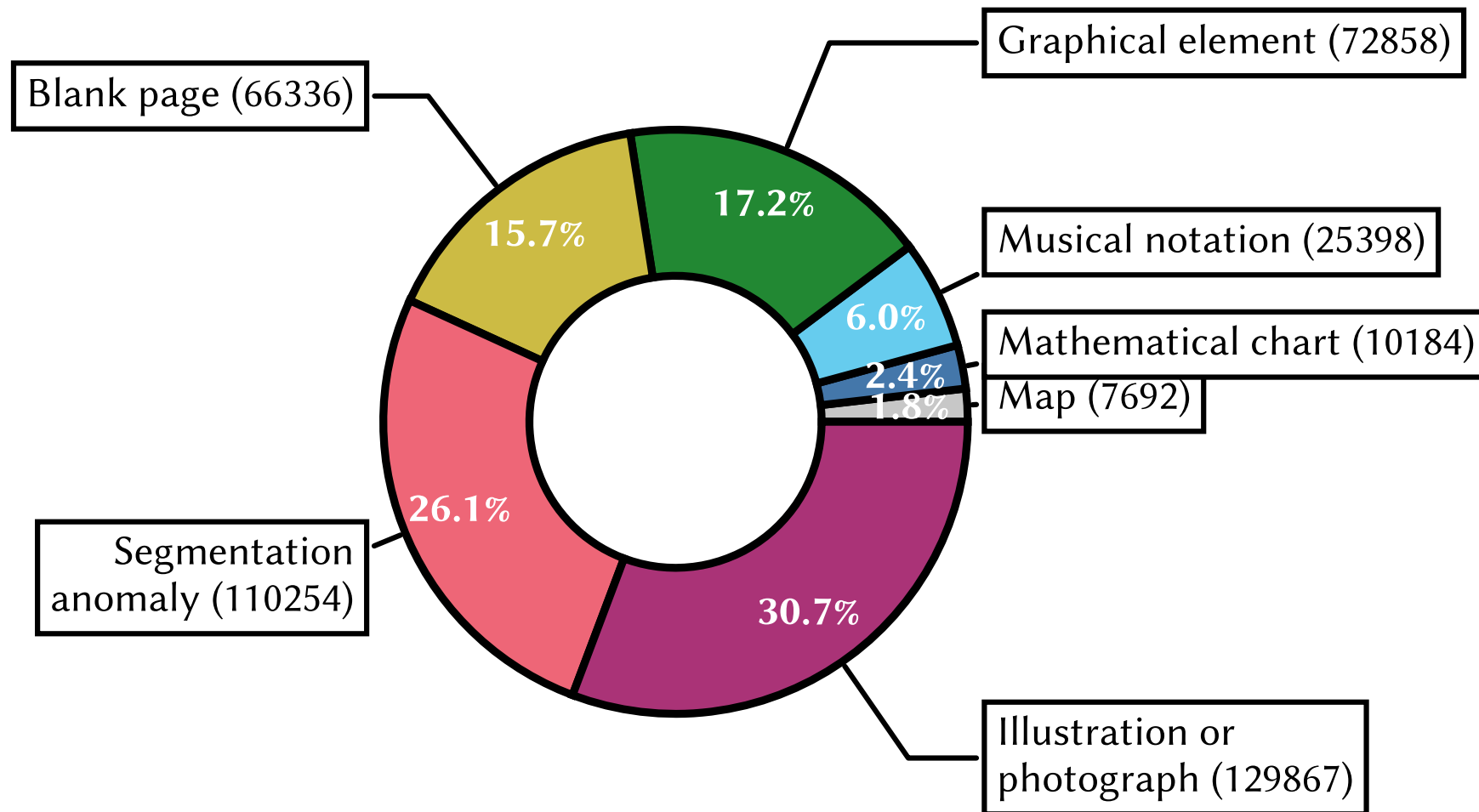
7 stamps

* Vi ser vekk fra segmenteringsfeil

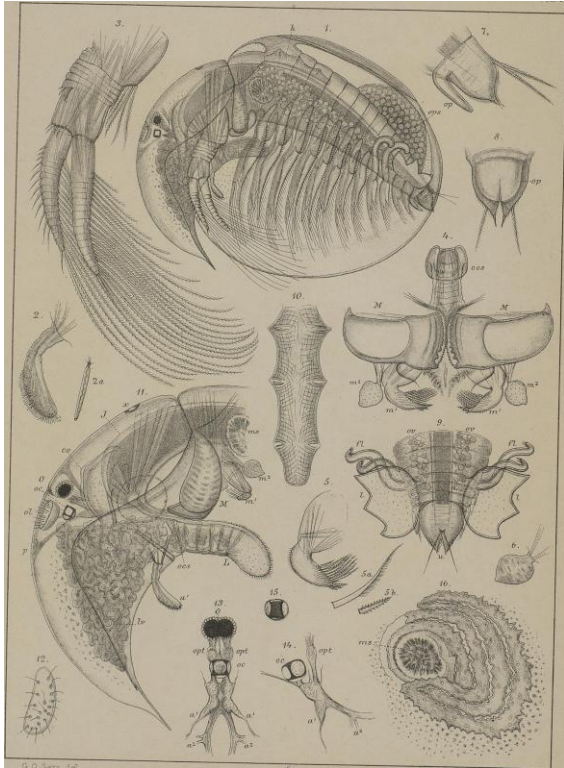
Alle* «falske negative» blanke sider
inneholder gjennomslag arkets bakside



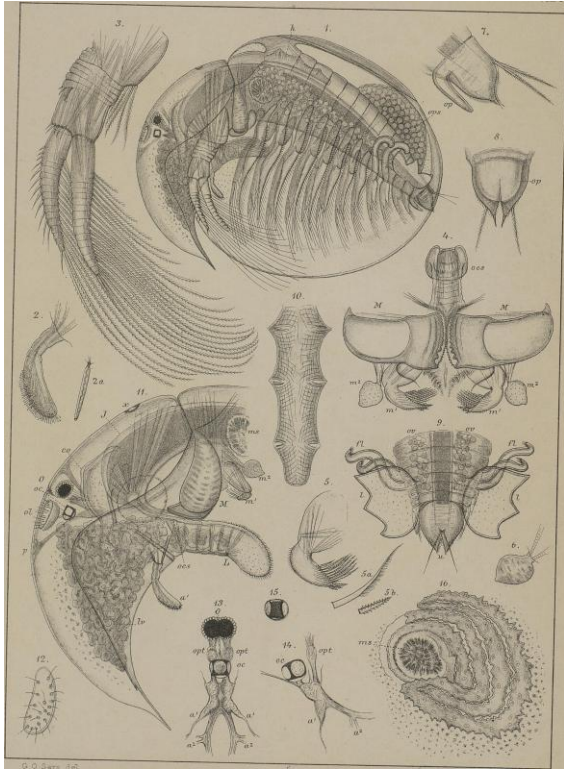
Vi brukte klassifiseringsmodellen til å estimere den totale klassefordelingen



Vi evaluerte modellenes evne til å gjenfinne transformerte varianter av kjente bilder



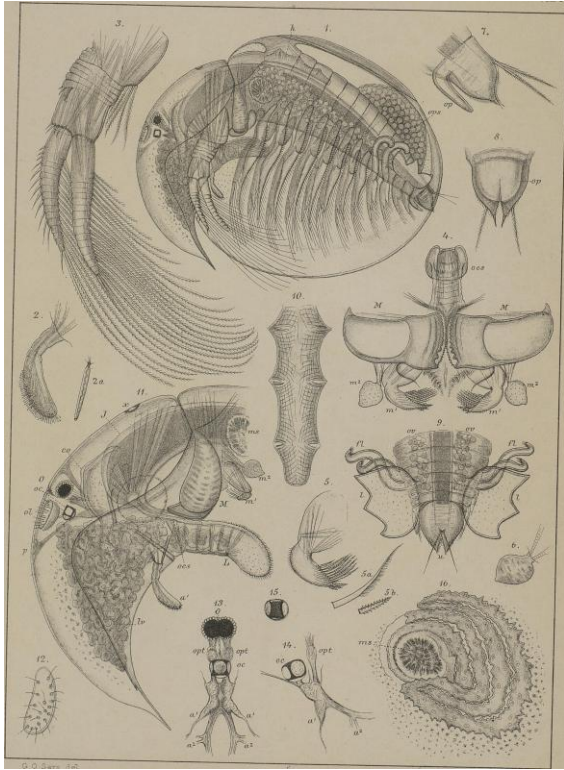
Vi evaluerte modellenes evne til å gjenfinne transformerte varianter av kjente bilder



Beskjer, skaler
og roter



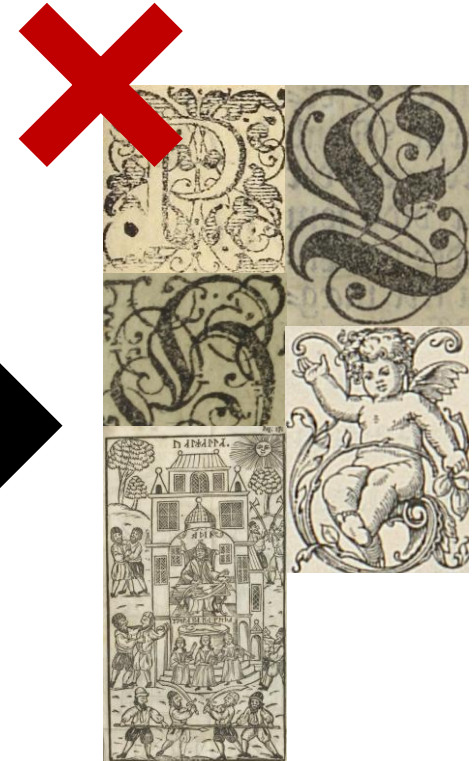
Vi evaluerte modellenes evne til å gjenfinne transformerte varianter av kjente bilder



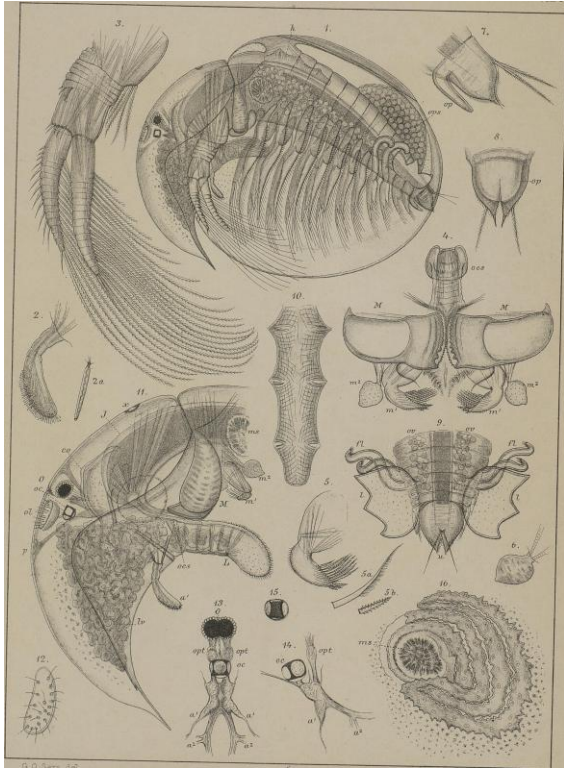
Beskjer, skaler
og roter



ViT-søk



Vi evaluerte modellenes evne til å gjenfinne transformerte varianter av kjente bilder



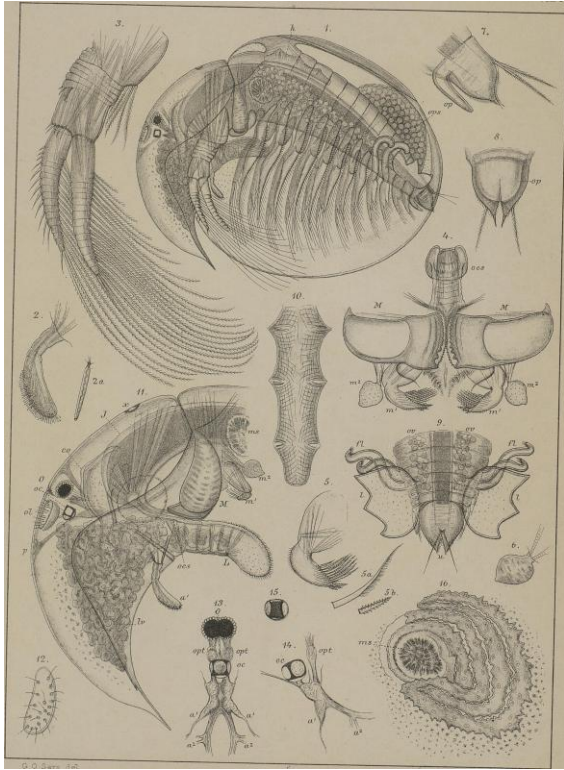
Beskjer, skaler
og roter



CLIP-søk



Vi evaluerte modellenes evne til å gjenfinne transformerte varianter av kjente bilder



Beskjer, skaler
og roter



SigLIP-søk



SigLIP modellen ga systematisk høyest Top-N nøyaktighet for evalueringssettet (684 bilder)

Accuracy Model	Top 1	Top 5	Top 10	Top 50
CLIP	492 (72 %)	596 (87 %)	613 (90 %)	638 (93 %)
SigLIP	529 (77 %)	633 (93 %)	645 (94 %)	665 (97 %)
ViT	529 (77 %)	582 (85 %)	597 (87 %)	612 (89 %)

The correct image was in the top 5 results for 93 % of the examples

Det er mange mulige utvidelser av
bildesøkappen

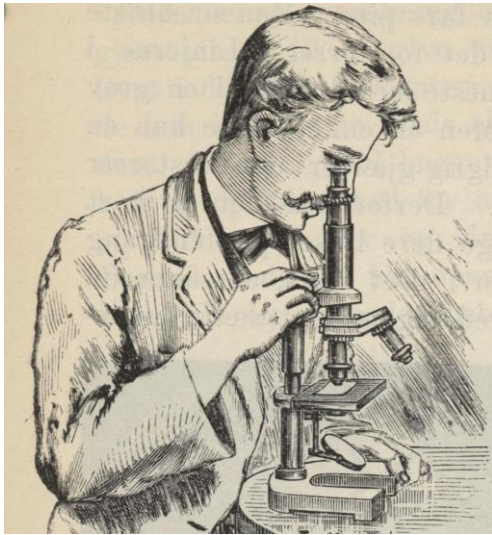


Domenespesifikke
bildetransformasjoner

Det er mange mulige utvidelser av bildesøkappen



Domenespesifikke
bildetransformasjoner

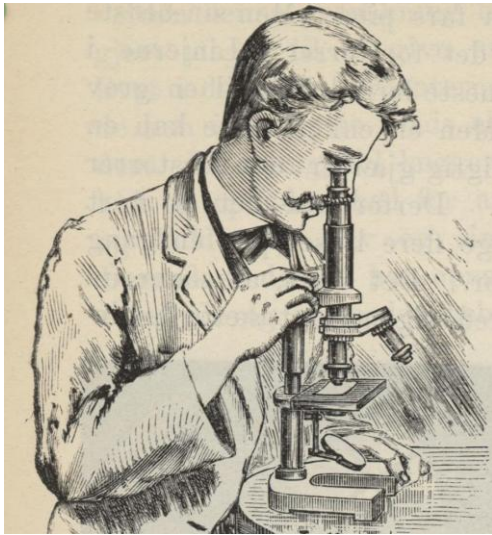


Grundigere evaluering
av søkekvalitet

Det er mange mulige utvidelser av bildesøkappen



Domenespesifikke
bildetransformasjoner



Grundigere evaluering
av søkekvalitet

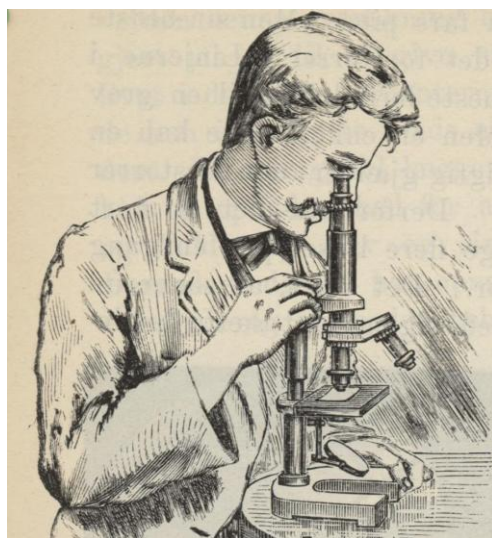


Hva «ser» modellen?

Det er mange mulige utvidelser av bildesøkappen



Domenespesifikke
bildetransformasjoner



Grundigere evaluering
av søkekvalitet



Hva «ser» modellen?




for dig. dyp dig
ned, op! se saa,
hu ren og pen og
ør med bær.
n nu mor!
e nu her!
g i mod!

e si-le fi-le æ-re
æ-de sæ-de y-de




mig faa en ka-ge!
kan faa den sø-de.
Tillatelse
Dette materiale er beskyttet
av opphevsrett. Les mer om
hva det betyr.

16



a-re va-re ra-re.
her ser jeg en ha-re.
aa-re daa-re taa-re.
to-re laa paa baa-re.

e-ne re-ne se-ne. mo-der he-der le-ne.
æ-re væ-re læ-re. hun gav mig en pæ-re.
li-de bi-de vi-de. o-le han kan ri-de.
ei-e vei-e fei-e.
ka-ren aas kan
nei-e.
øi-e mõi-e tõi-e.
bog-en læ-ses
nøi-e.



u-ren mu-ren tu-ren, der er lyd i lu-ren.
ø-re hø-re fø-re. luk nu di-ne dø-re!
u-de ru-de tu-de, sov nu paa din pu-de!

17

e-len da-len fei-er sa-len.
ka-ren aa-len ko-ger kaa-len.

9.

al-le bal-le am-me tam-me an-na
han-na op-pe top-pe yp-pe dyp-pe.
rok-ken sur-rer. kat-ten mur-rer.
or-ren kur-rer. kol-len rul-ler.
buk-ken hop-per bæk-ken.
gut-ten let-ter sæk-ken.
gun-nar fos-sen ret-ter ryg-gen,
sæt-ter hat-ten paa sin nak-ke.



beg-ge dis-se kat-te
faar sig mad i kop-
pen; men mo-der
kat-te vat-te tog en
rot-te i en pot-te.

lot-te læg-ger du
ne lom-mer er ik
to pak-ker med s
ker og pik-ker m

er-le per-le
ev-ne hev-ne
av-le tav-le
il-de sil-de
en-ge sen-ge



ug-
kom-
el-v
syn-
ja-g
den har-ver. ko
dyr-kes. en-gen v
man-den hug-ger.
vi har en ko, s
gaar u-de paa

<https://dh.nb.no/run/bildesok/>

Kort fortalt kan KI-modeller hjelpe oss å utforske
Nasjonalbibliotekets bøker basert bildene i dem



<https://github.com/Sprakbanken/CHR24-image-retrieval>

