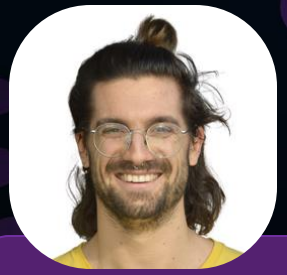
CoMix: A comprehensive Benchmark for Multi-Task Comic Understanding

NEURAL INFORMATION PROCESSING SYSTEMS



Emanuele Vivoli

Marco Bertini

Dimosthenis Karatzas

CVC Computer Vision Center

Computer Vision Center - UAB, Barcelona

MICC - University of Florence, Italy

GitHub

Comics Datasets

d: Detection
c: Classification
t2c: Text-to-Character
c2c: Character-to-Character
N: Character naming
D: Dialog generation

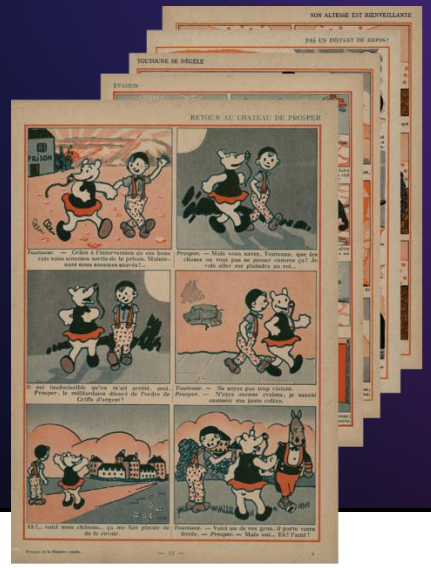
Dataset	Release	Avail	Tasks	Years	Style	Books	Pages
eBDtheque	2013	✓	d,t2c	1905-2012	mix	28	100
COMICS	2017	✓	c	1938-1954	comics	3948	198k
GCN	2017	✗	d,t2c	1978-2013	comics	*253	*38k
DCM772	2018	✓	d	1938-1954	comics	27	772
Manga109	2018	✓	d,t2c,c2c	1970-2010	manga	109	10k
BCBId	2022	✓	-	-	bangla	64	3k
VLRC	2023	✗	-	1940-now	-	*376	*7k
PopManga	2024	✓	d,t2c,c2c	2010-2023	manga	25	1.8k
CoMix (our)	2024	✓	d,t2c,c2c,N,D	1938-2023	mix	100	3.8k

Data

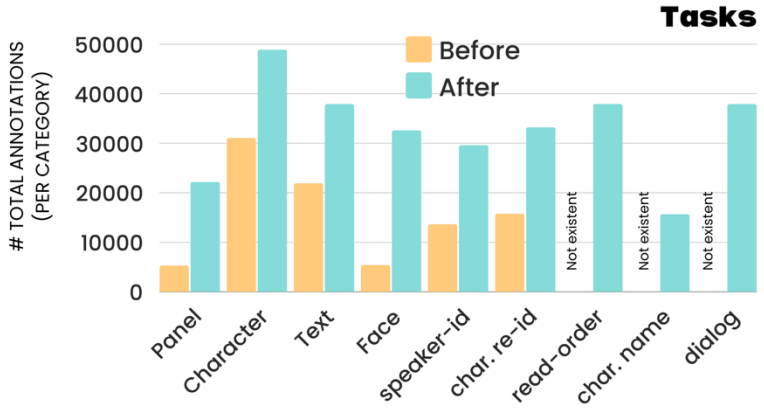
Comics

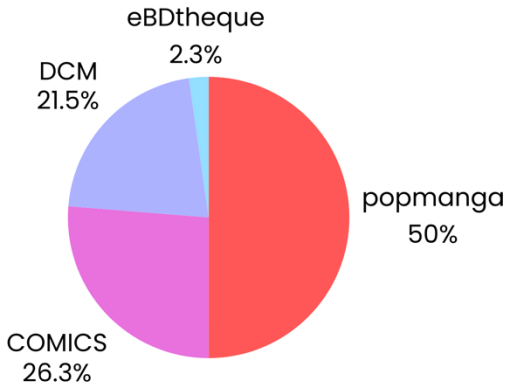
DCM

eBDtheque



PopManga





Annotations

Detection problems:

- Missing characters

BEFORE

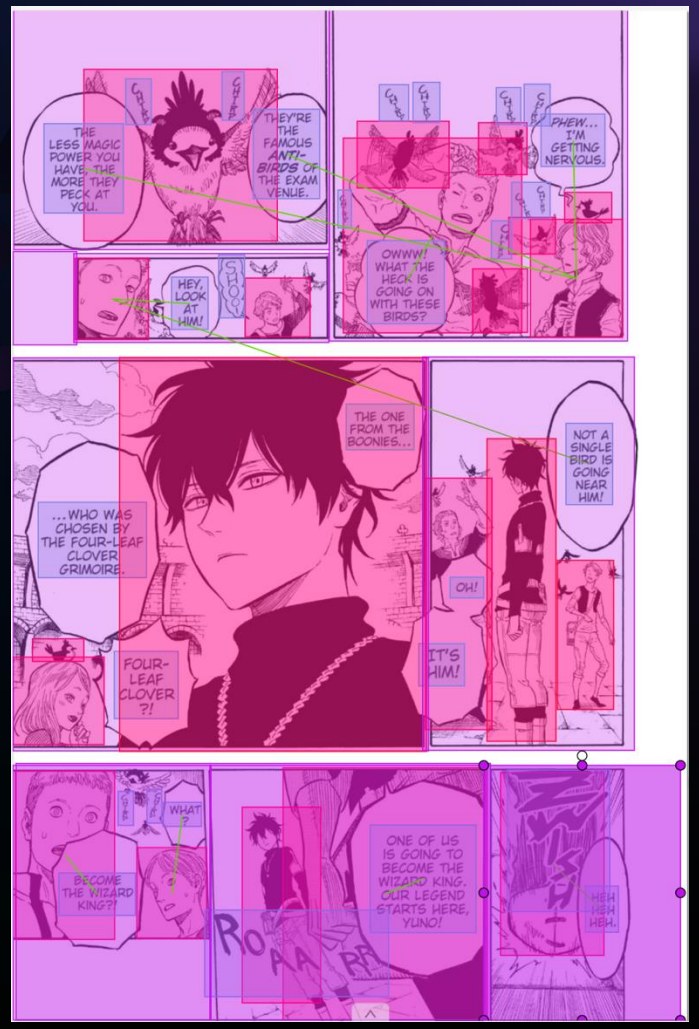
AFTER

Annotations

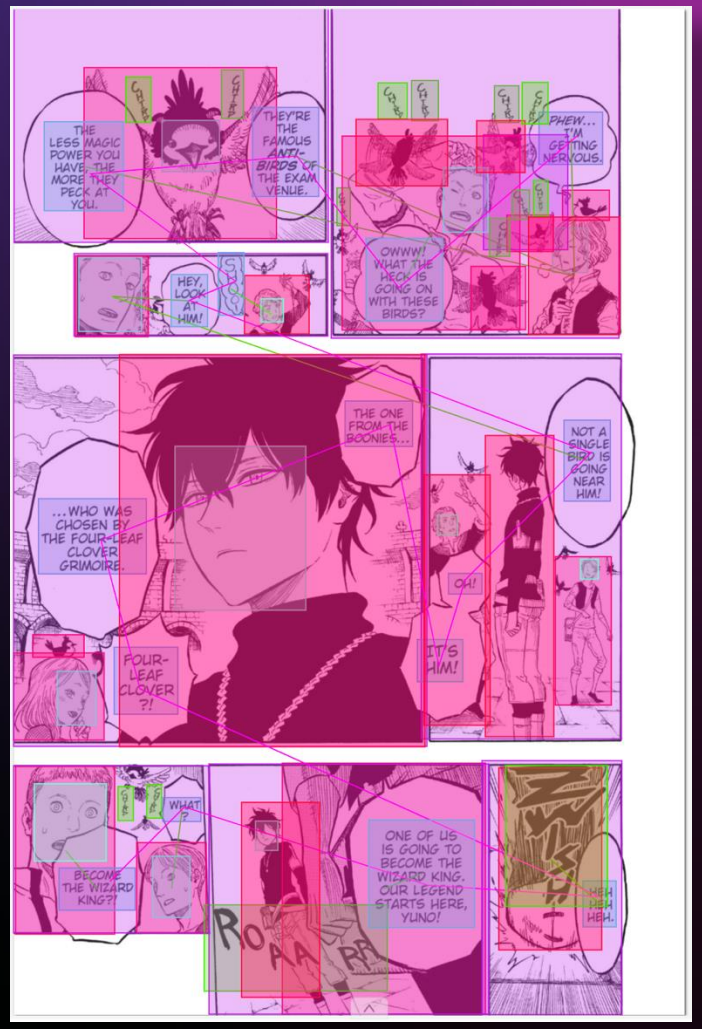
Detection problems:

- Missing characters
- Not precise bboxes

BEFORE



AFTER

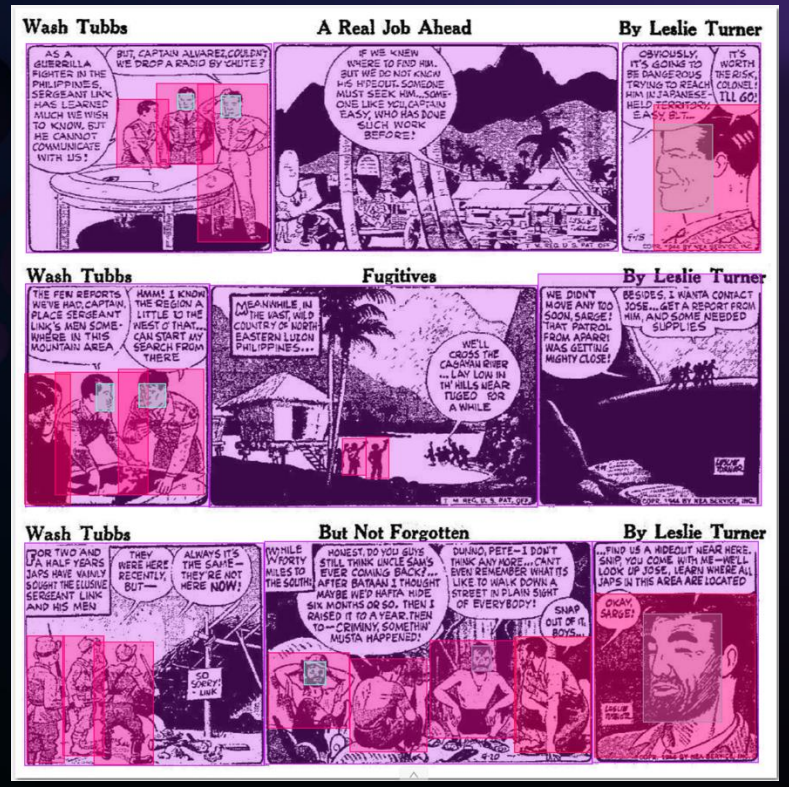


Annotations

Detection problems:

- Missing characters
- Not precise bboxes
- Not uniform annotations

BEFORE



AFTER

Annotations

Detection problems:
- Missing characters
- Not precise bboxes
- Not uniform annotations

Not existing Tasks:
- Character Naming
- Dialog generation

① **Uncle Sam:** "GRR!! I SNAP YOUR NECK LIKE PIGEON!"
② **Char 1:** "NO! GASPS NO! DON'T! AGGG-GG!"
③ **Iron Ace:** "WELL, THERE'S A MAN WHO DOES A GOOD [...]"
④ **Iron Ace:** "I'M GETTING MY USUAL GOOD-BYE... I DON'T [...]"
Iron Ace: "NOW.. IF THEY ONLY KEEP DOING THE [...]"
Radio Operator: "PATROL PLANES ATTENTION... [...]"
⑦ **Iron Ace:** "OH-OH!! THE WHOLE JAP AIR FORCE IS [...]"
⑧ **Iron Ace:** "WELL, COME AND GET IT, BOYS.. BUT IT'S THE [...]"

Tasks

Object Detection

Speaker id.

Character Re-Id

Reading order

Character Naming

Narrator, **Sailor 1**, **McWhustle**, **Captain**
Matey, Sailor 2, **Sailor 3**, Sailor 4

Dialog generation

① **Narrator:** "As the tramp steamer SS. Clementine crosses the Equator [...]"
② **Sailor 1:** "THAT DISGUISE DOESN'T FOOL ME! YOU'RE CHIEF ENGINEER [...]"
③ **McWhustle:** "HOOT, LADDIE! BEFORE YE CAN BE A SON OF NEPTUNE, YE [...]"
④ **McWhustle:** "HOOT, MON! WE'RE AGROUND ON A REEF!"
Captain: "THERE ARE NO REEFS IN THESE PARTS! BACK TO YOUR ENGINE [...]"
Captain: "WHAT'S WRONG WITH YOUR BLASTED ENGINES, MCWHUSTLE ?"
McWhustle: "WHY DON'T YE LEARN TO NAVIGATE YER SHIP, LASSITER! [...]"
McWhustle: "NO HARD FEELINGS, CAP'N! BUT THE ENGINES ARE DOING FINE!"
Captain: "I KNOW THAT, MAC... WE'VE BEEN SHIPMATES TOO LONG TO [...]"
Captain: "WE SHOULD BE MAKING HEADWAY... BUT SOMETHING'S [...]"
McWhustle: "MON, IT'S NO CANNY!"
Narrator: "Meanwhile, below..."
Matey: "THE SKIPPER' SAYS WE CAN'T HAVE RUN AGROUND!"
⑭ Sailor 2: "HEY, MATEY! LOOK!"
⑮ **Sailor 3**: "AHOY, TOPSIDES! STAND BY TO REPEL BOARDERS IN THE [...]"
⑯ Sailor 4: "BOARDERS IN THE STOKEHOLD? I NEVER HEARD OF SUCH A THING."

Tasks

Object Detection

mAP – R@k

Panels Characters Textboxes Faces

Speaker id.

R@#text

Character Re-Id

AMI - NMI

Reading order

Edit distance

Character Naming

ANLS

Dialog generation

Hybrid Dialog Score

Algorithm 2 Hybrid Dialog Score

```
procedure EvaluateTranscription(model_output, ground_truth)
    matches ← find optimal matches(model_output, ground_truth)
    tot_ed, char_name_score ← 0, 0, 0
    for each (mo, gt) in matches do
        edit_dist ← calculate edit distance(mo.text, gt.text)
        tot_ed ← tot_ed + edit_dist / len(gt.text)
        anls_score ← calculate ANLS(mo.name, gt.name)
        char_name_score ← char_name_score + anls_score
    end for
    tot_ed ← 1− tot_ed
    char_name_score ← char_name_score / len(matches)
    return tot_ed, char_name_score
end procedure
```

Hybrid Dialog Score

Ground Truth

1. **Uncle Sam:** "GRR!! I SNAP YOUR NECK LIKE PIGEON!"
2. **Char 1:** "NO! GASPS NO! DON'T! AGGG-GG!"
3. **Iron Ace:** "WELL, THERE'S A MAN WHO DOES A GOOD [...]"
4. **Iron Ace:** "I'M GETTING MY USUAL GOOD-BYE... I DON'T [...]"
5. **Iron Ace:** "NOW.. IF THEY ONLY KEEP DOING THE [...]"
6. **Radio Operator:** "PATROL PLANES ATTENTION... [...]"
7. **Iron Ace:** "OH-OH!! THE WHOLE JAP AIR FORCE IS [...]"
8. **Iron Ace:** "WELL, COME AND GET IT, BOYS.. BUT IT'S THE [...]"

GPT4

1. **US Soldier:** "GRR!! I SNAP YOUR NECK LIKE PIGEON!"
2. **Pilot 1:** "NO! GASPS NO! DON'T! AGGG-GG!"
3. **Pilot 1:** "WELL, THERE'S A MAN WHO DOES A GOOD [...]"
4. **Pilot 1:** "NOW.. IF THEY ONLY KEEP DOING THE MISSING [...]"
5. **Pilot 1:** "OH-OH!! THE WHOLE JAP AIR FORCE IS AFTER [...]"
6. **Pilot 1:** "I'M GETTING MY USUAL GOOD-BYE... I DON'T [...]"
7. **Radio Operator:** "PATROL PLANES ATTENTION... [...]"
8. **Pilot 1:** "WELL, COME AND GET IT, BOYS... BUT IT'S THE [...]"

Hybrid Dialog Score

Algorithm 2 Hybrid Dialog Score

```
procedure EVALUATETRANSCRIPTION(model_output, ground_truth)
    matches ← find optimal matches(model_output, ground_truth)
    tot_ed, char_name_score ← 0, 0, 0
    for each (mo, gt) in matches do
        edit_dist ← calculate edit distance(mo.text, gt.text)
        tot_ed ← tot_ed + edit_dist / len(gt.text)
        anls_score ← calculate ANLS(mo.name, gt.name)
        char_name_score ← char_name_score + anls_score
    end for
    tot_ed ← 1− tot_ed
    char_name_score ← char_name_score / len(matches)
    return tot_ed, char_name_score
end procedure
```

Ground Truth

① **Uncle Sam:** "GRR!! I SNAP YOUR NECK LIKE PIGEON!"
② **Char 1:** "NO! GASPS NO! DON'T! AGGG-GG!"
③ **Iron Ace:** "WELL, THERE'S A MAN WHO DOES A GOOD [...]"
④ **Iron Ace:** "I'M GETTING MY USUAL GOOD-BYE... I DON'T [...]"
Iron Ace: "NOW.. IF THEY ONLY KEEP DOING THE [...]"
Radio Operator: "PATROL PLANES ATTENTION... [...]"
⑦ **Iron Ace:** "OH-OH!! THE WHOLE JAP AIR FORCE IS [...]"
⑧ **Iron Ace:** "WELL, COME AND GET IT, BOYS.. BUT IT'S THE [...]"

GPT4

① **US Soldier:** "GRR!! I SNAP YOUR NECK LIKE PIGEON!"
② **Pilot 1:** "NO! GASPS NO! DON'T! AGGG-GG!"
③ **Pilot 1:** "WELL, THERE'S A MAN WHO DOES A GOOD [...]"
④ **Pilot 1:** "NOW.. IF THEY ONLY KEEP DOING THE MISSING [...]"
Pilot 1: "OH-OH!! THE WHOLE JAP AIR FORCE IS AFTER [...]"
Pilot 1: "I'M GETTING MY USUAL GOOD-BYE... I DON'T [...]"
⑦ **Radio Operator:** "PATROL PLANES ATTENTION... [...]"
⑧ **Pilot 1:** "WELL, COME AND GET IT, BOYS... BUT IT'S THE [...]"

sentences

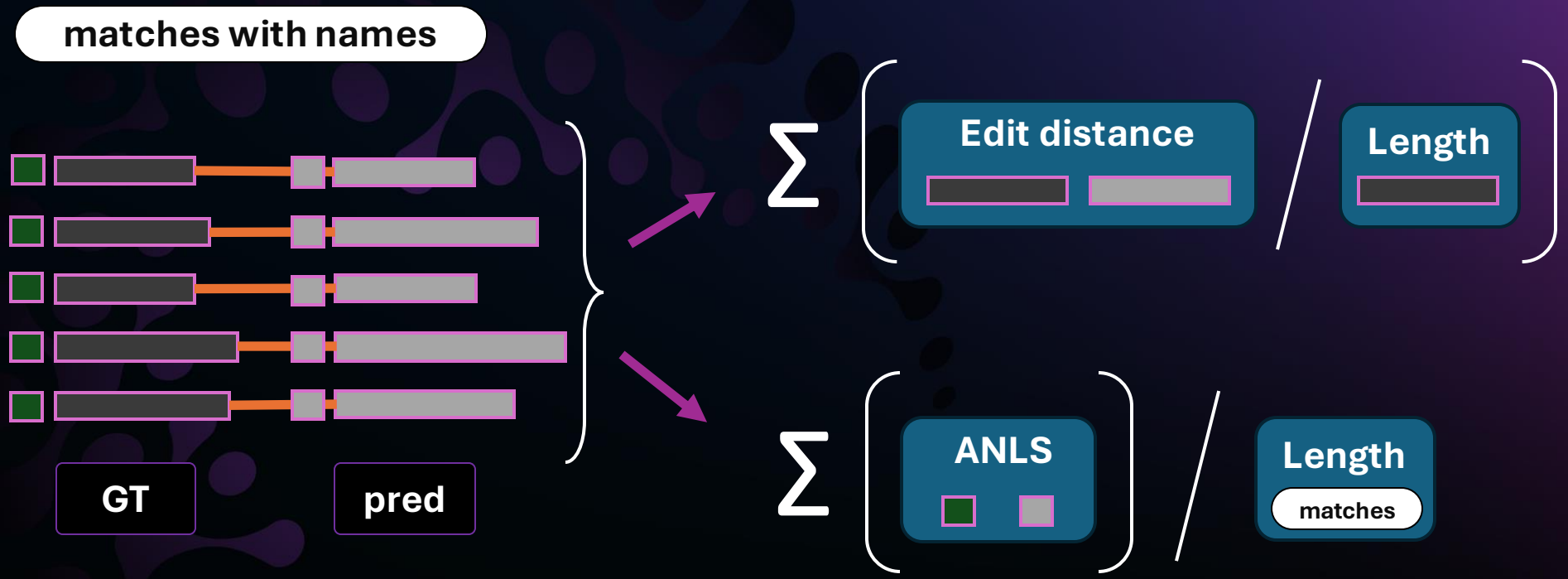
Hungarian Matching with **Edit distance**

matches

Hybrid Dialog Score

Algorithm 2 Hybrid Dialog Score

```
procedure EVALUATETRANSCRIPTION(model_output, ground_truth)
    matches ← find optimal matches(model_output, ground_truth)
    tot_ed, char_name_score ← 0, 0, 0
    for each (mo, gt) in matches do
        edit_dist ← calculate edit distance(mo.text, gt.text)
        tot_ed ← tot_ed + edit_dist / len(gt.text)
        anls_score ← calculate ANLS(mo.name, gt.name)
        char_name_score ← char_name_score + anls_score
    end for
    tot_ed ← 1− tot_ed
    char_name_score ← char_name_score / len(matches)
    return tot_ed, char_name_score
end procedure
```



Benchmarks

Detection:
- YOLO, SSD, FasterRCNN
- Magi, Gdino

Speaker-id:
- heuristic, Magi

Character Re-Id:
- CLIP, DINO, Magi

Character Naming:
- GPT4

Dialog generation:
- Magi, GPT4

Task	Output	Metric	Baseline	Score
Object detection	box detection	mAP - R@100	Magi	78.6 - 67.9
Speaker identification	object indexes	R@#text	heuristic	0.68
Character Re-Id	cluster ids	AMI - NMI	DINOv2	0.29 - 0.51
Character Naming	names	ANLS	GPT-4	47.11
Dialog generation	list of tuples	HDS	GPT-4	93.14

Thanks for your attention

NEURAL INFORMATION PROCESSING SYSTEMS

CoMix Repo

Our Survey

Poster session 5

Fri 13 Dec
11 a.m. PST

micc

CVC

Computer Vision Center