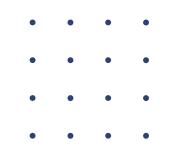
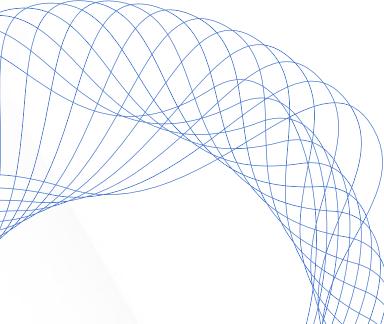


RESEARCH PROJECT

Kanokporn Pakdeenual

675020069-9



TITLE

Published In



Computer Vision – ECCV
2024: 18th European
Conference, Milan, Italy,
September 29 – October 4,
2024, Proceedings, Part
XVIII

Sep 2024 | 578 pages

ISBN: 978-3-031-72648-4

DOI: 10.1007/978-3-031-72649-1

Editors: Aleš Leonardis, + 5

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025.

Publisher

Springer-Verlag
Berlin, Heidelberg

Publication History

Published: 30 September 2024

Zero-Shot Detection of AI-Generated Images

Davide Cozzolino¹ , Giovanni Poggi¹ , Matthias Nießner² , and Luisa Verdoliva^{1,2}

¹ University Federico II of Naples, 80125 Naples, Italy

² Technical University of Munich, 85748 Garching, Germany

{davide.cozzolino,poggi,verdoliv}@unina.it, niessner@tum.de

Google Scholar

Zero-Shot Detection of AI-Generated Images



บทความ

ผลการค้นหาประมาณ 5,940 รายการ (0.28 วินาที)

ເວລາໄດ້ໃຫ້
ຕັ້ງແຕ່ 2025
ຕັ້ງແຕ່ 2024
ຕັ້ງແຕ່ 2021
ກໍາທັນດຂ່ວງເອງ...

Zero-shot detection of ai-generated images

D Cozzolino, G Poggi, M Nießner... - European Conference on ..., 2024 - Springer

... Full size **image** To this end, we propose a **zero-shot detection** method that only requires real **images** for learning their underlying distribution. Our key idea is to use lossless coding and ...

☆ ບັນທຶກ ۷۹ อ້າງອີງ อ້າງໂດຍ 19 ນທຄວາມທີ່ເກີຍວ່າຂອງ ທັ້ງໝົດ 9 ລັບນັ້ນ ☺



Conference Ranks

บทความ “Zero-Shot Detection of AI-Generated Images” เป็นงานนำเสนอในการประชุม ECCV 2024 (European Conference on Computer Vision) ซึ่งถูกตีพิมพ์ใน Proceedings (Lecture Notes in Computer Science, LNCS Vol. 15076) จาก Springer-Verlag

Conference Data				
Name	↓	Abrv.	↑ Rank	↑ Source ↑
European Conference on Computer Vision		ECCV	A	ERA
European Conference on Computer Vision		ECCV	A1	Qualis
Name		Abrv.	Rank	Source

Zero-Shot Detection of AI-Generated Images

Davide Cozzolino¹ , Giovanni Poggi¹ , Matthias Nießner² , and Luisa Verdoliva^{1,2} 

¹ University Federico II of Naples, 80125 Naples, Italy

² Technical University of Munich, 85748 Garching, Germany

{davide.cozzolino,poggi,verdoliv}@unina.it, niessner@tum.de

Abstract. Detecting AI-generated images has become an extraordinarily difficult challenge as new generative architectures emerge on a daily basis with more and more capabilities and unprecedented realism. New versions of many commercial tools, such as DALL·E, Midjourney, and Stable Diffusion, have been released recently, and it is impractical to continually update and retrain supervised forensic detectors to handle such a large variety of models. To address this challenge, we propose a zero-shot entropy-based detector (ZED) that neither needs AI-generated training data nor relies on knowledge of generative architectures to artificially synthesize their artifacts. Inspired by recent works on machine-generated text detection, our idea is to measure how *surprising* the image under analysis is compared to a model of real images. To this end, we rely on a lossless image encoder that estimates the probability distribution of each pixel given its context. To ensure computational efficiency, the encoder has a multi-resolution architecture and contexts comprise mostly pixels of the lower-resolution version of the image. Since only real images are needed to learn the model, the detector is independent of generator architectures and synthetic training data. Using a single discriminative feature, the proposed detector achieves state-of-the-art performance. On a wide variety of generative models it achieves an average improvement of more than 3% over the SoTA in terms of accuracy. Code is available at <https://grip-unina.github.io/ZED/>.

Abstract



Abstract

1. Problem – ໃຄງກາ

- Nowadays, there are new AI image generation tools (Generative Models) such as DALL·E, Midjourney, and Stable Diffusion that are being rapidly developed.
- Traditional forensic detectors are no longer sufficient to detect images generated by AI, as they were trained on older patterns.
- However, new models keep evolving → causing delays and poor detection performance.
- Therefore, it is necessary to find methods that do not require image detection models or fake images for training.

Abstract

2. Proposed Solution – ວິທີກາຣົກ່າວເສນວ

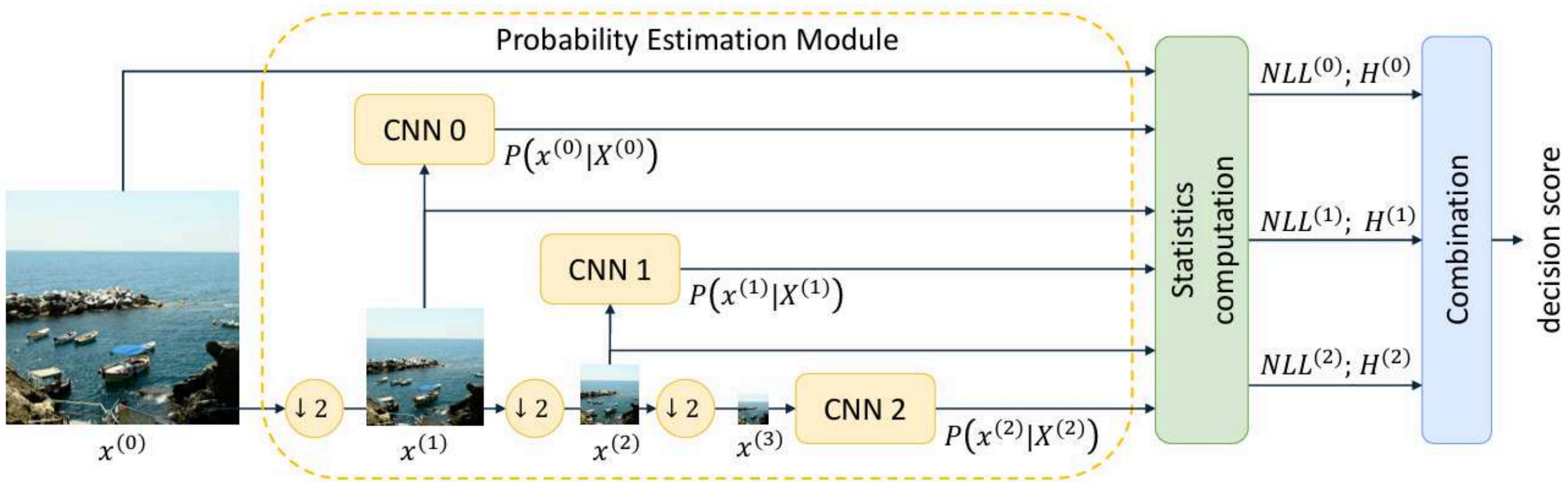
Proposed Method: ZED (Zero-shot Entropy-based Detector)

- Propose a forensic detection tool called ZED, which operates in a "Zero-Shot" manner.
- Key advantage: Does not require any AI-generated image samples, not even a single one.
- Enables detection of images from newly developed AI models immediately, without needing prior examples.

Related work

กลุ่มแนวทาง	ใช้ภาพ จริงในการ เทรน	ใช้ภาพ ปลอมใน การเทรน	วิธีเทรนโดยสรุป	จุดเด่น	จุดด้อย
1. Supervised Learning	ใช้	ใช้ (จำนวน มาก)	ฝึกโมเดลให้จำแนกภาพ “จริง” กับ “ปลอม” โดยใช้ ข้อมูลทั้งสองแบบที่มีป้าย กำกับ (เช่น 0 = จริง, 1 = ปลอม)	- แม่นยำสูงถ้าเจอกาแฟปลอม แบบเดียวกับที่ฝึก - ใช้ง่าย เพราะเป็นการจำแนก สองกลุ่ม	- ต้องใช้ข้อมูลจำนวนมาก - ถ้าเจอ AI รุ่นใหม่ที่ไม่เคยเห็น → จำแนกไม่ออก
2. Few-shot / Incremental Learning	ใช้	ใช้ (จำนวน น้อย)	เริ่มจากโมเดลที่มือyuแล้ว แล้ว ฝึกเพิ่มเล็กน้อยด้วยภาพ ปลอมรุ่นใหม่ (fine-tune หรือ update โมเดลเรื่อย ๆ)	- ใช้ข้อมูลปลอมน้อยลง - ปรับตัวตาม generator ใหม่ ได้เร็ว	- ปรับตัวตาม generator ใหม่ ได้เร็ว - ยังต้องใช้ภาพปลอม - ถ้า generator ใหม่ต่างจาก เดิมมาก → ผลตก
3. Zero-shot Learning (ZED)	ใช้	ไม่ใช้เลย	ฝึกด้วยภาพจริงล้วน ให้โมเดล เรียนรู้ “ธรรมชาติของภาพ จริง” จากนั้นคำนวณ NLL กับ Entropy ของภาพใหม่ เพื่อดูว่ามัน “เข้ากับโลกจริง ไหม”	- ไม่ต้องใช้ภาพปลอมเลย - ตรวจจับได้กับ AI รุ่นใหม่ทุก แบบ	- ต้องออกแบบสติ๊กเกอร์ (NLL-H) - พึงความแม่นของโมเดลภาพ จริงสูง

Method



Method

เพื่อที่จะสร้าง "แบบจำลองของภาพจริง" เปเปอร์นี้ได้เลือกใช้เครื่องมือที่เรียกว่า **Lossless Image Encoder** เป็นเครื่องมือที่พยายามบีบอัดภาพจริงโดยทำนายพิกเซลแบบไม่สูญเสียข้อมูล ผ่านการคาดคะเนจากภาพระดับต่ำขึ้นมาสู่ระดับสูง ซึ่งในที่นี่คือโมเดลที่ชื่อว่า **SReC (Super-Resolution based lossless Compressor)**

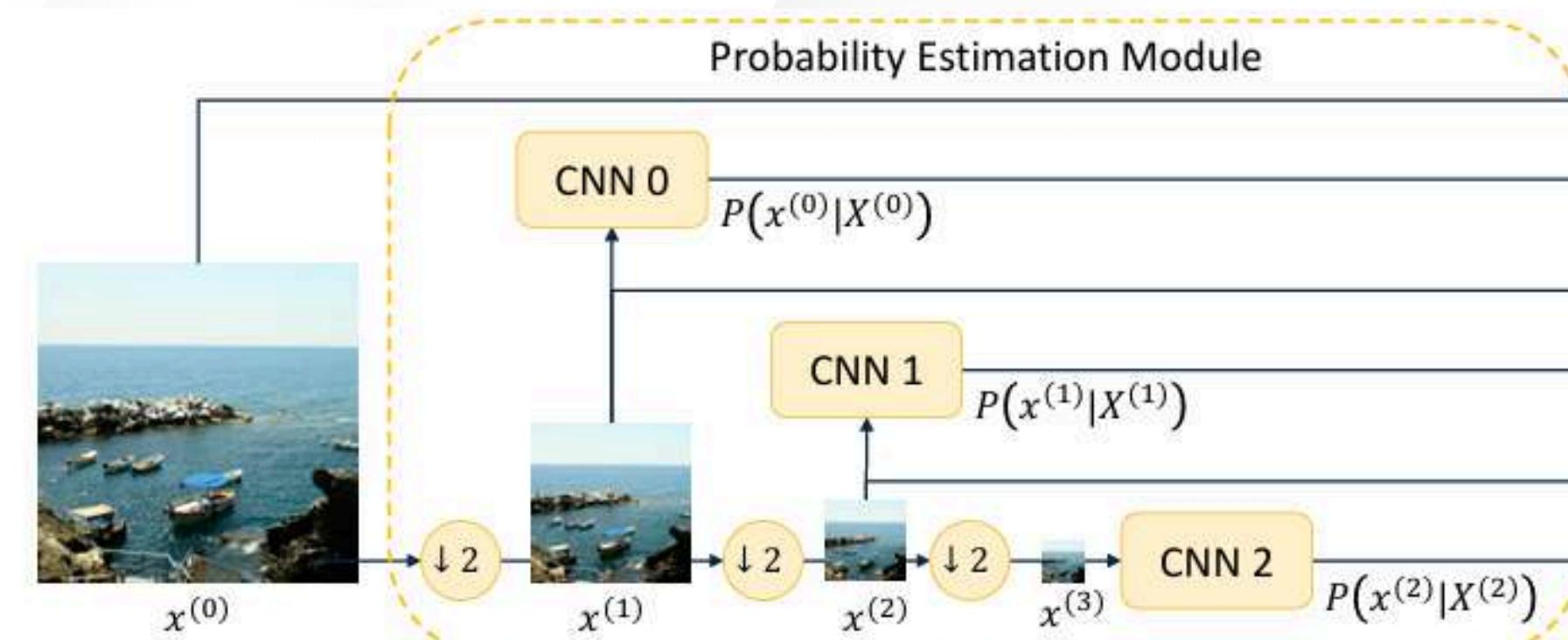
ถ้าภาพไหนโมเดล “บีบได้ไม่ดี” (ทำนายพิกเซลไม่เข้า) ภาพนั้นน่าจะไม่ใช่ของจริง

Method

SReC (Super-Resolution based lossless Compressor)

โครงสร้างโดยรวม

1. ทำภาพหลายความละเอียด เริ่มจากนำภาพจริงมาย่อขนาดลง 3 ระดับด้วยวิธี average pooling โดยบีบอัดภาพย่อลงที่ลดครึ่งของความกว้างและความสูง ทำให้เรามีภาพทั้งหมด 4 ความละเอียด ตั้งแต่เล็กสุดไปจนถึงขนาดเดิม

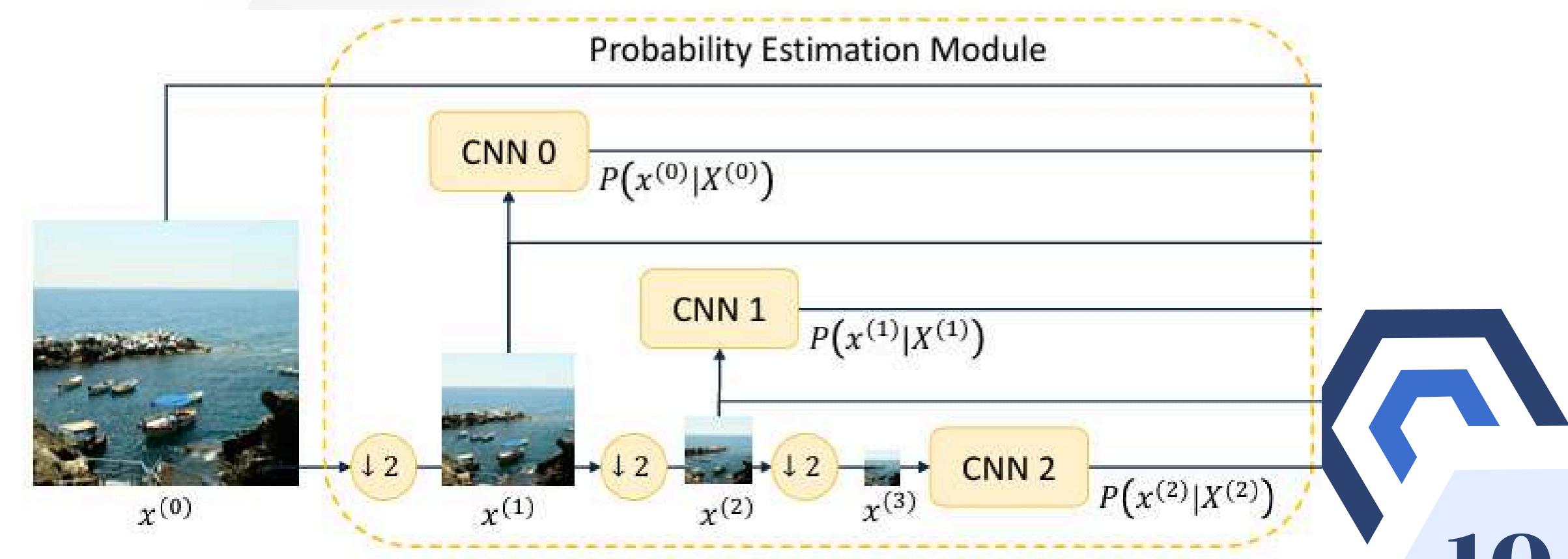


Method

SReC (Super-Resolution based lossless Compressor)

โครงสร้างโดยรวม

2. CNN ทำการเรียนรู้ความสัมพันธ์ของพิกเซลในภาพจริง โดยเข้าถึงโครงสร้างของรูปภาพและการแจกแจงความน่าจะเป็นของพิกเซลในแต่ละขนาดของภาพ เพื่อใช้คำนวณค่าพิกเซลถัดไปได้แม่นยำว่าโอกาสที่พิกเซลบริเวณนี้จะมีการแจกแจงของพิกเซลเป็นอย่างไรจากบริบทรอบๆ



Method

3. ค่า **Entropy (H)** และ **NLL (Negative Log-Likelihood)** ถูกใช้เพื่อวัดความยากในการเข้ารหัสภาพผ่านโมเดลบีบอัด โดยมีเป้าหมายเพื่อแยกแยะระหว่างภาพจริงและภาพปลอมผ่านความซับซ้อนและความสามารถในการคาดเดาของภาพแต่ละภาพ

Entropy (H) – เอนโทรปี

เป็นการวัดว่า "ข้อมูลนี้มีความคาดเดาได้ยากแค่ไหน" ถ้าภาพมีความซับซ้อนมาก ค่า Entropy จะ สูง ถ้าภาพเรียบง่าย ค่า Entropy จะ ต่ำ

NLL (Negative Log-Likelihood)

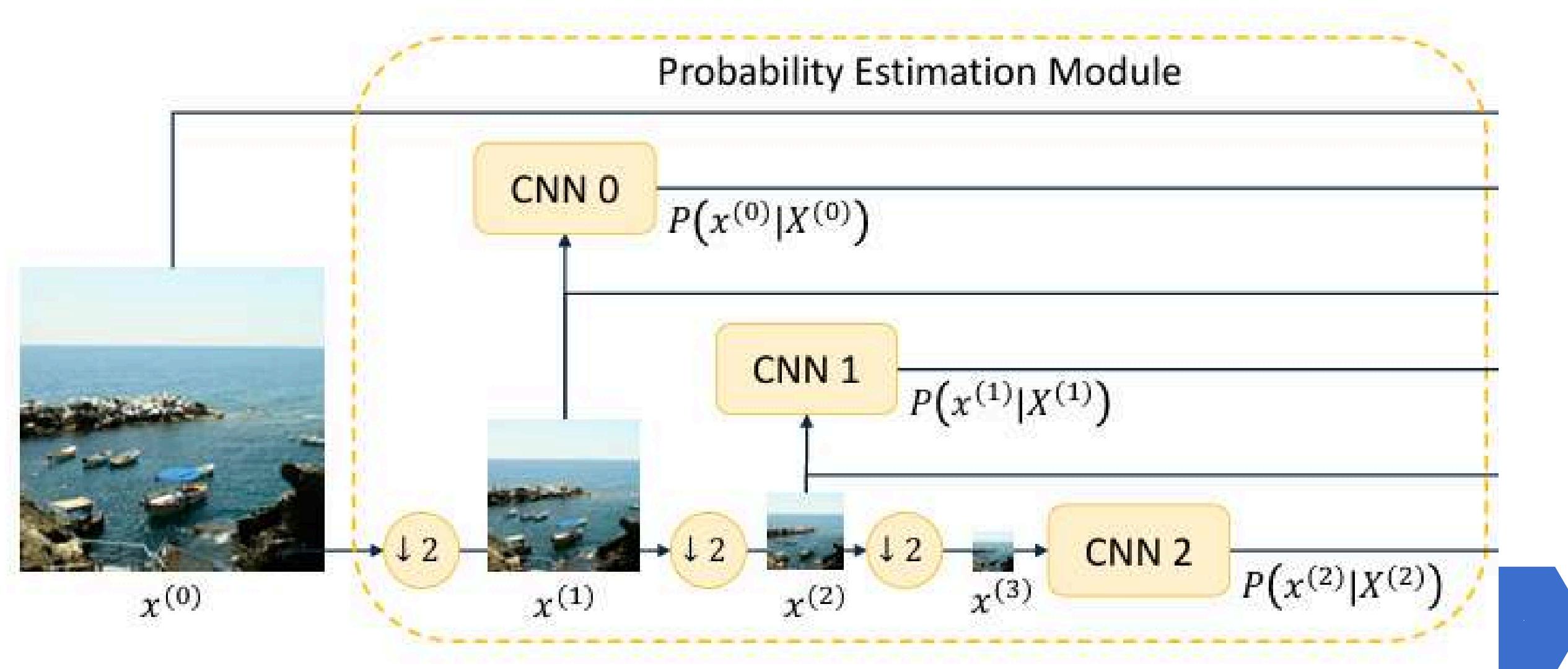
ต้นทุนการเข้ารหัส
เป็นการวัดว่า "ค่าความยากในการบีบอัดหรือคาดเดาภาพ" ถ้าโมเดลสามารถคาดเดาได้ดี ค่า NLL จะต่ำ ถ้าโมเดลคาดเดาไม่เก่ง ค่า NLL จะสูง

Method

4. กระบวนการทำงาน (ZED: Zero-shot Entropy-based Detector)

กระบวนการทำงานของวิธีที่นำเสนอ (เรียกว่า ZED) สามารถแบ่งเป็นขั้นตอนย่อยๆ ได้ดังนี้:

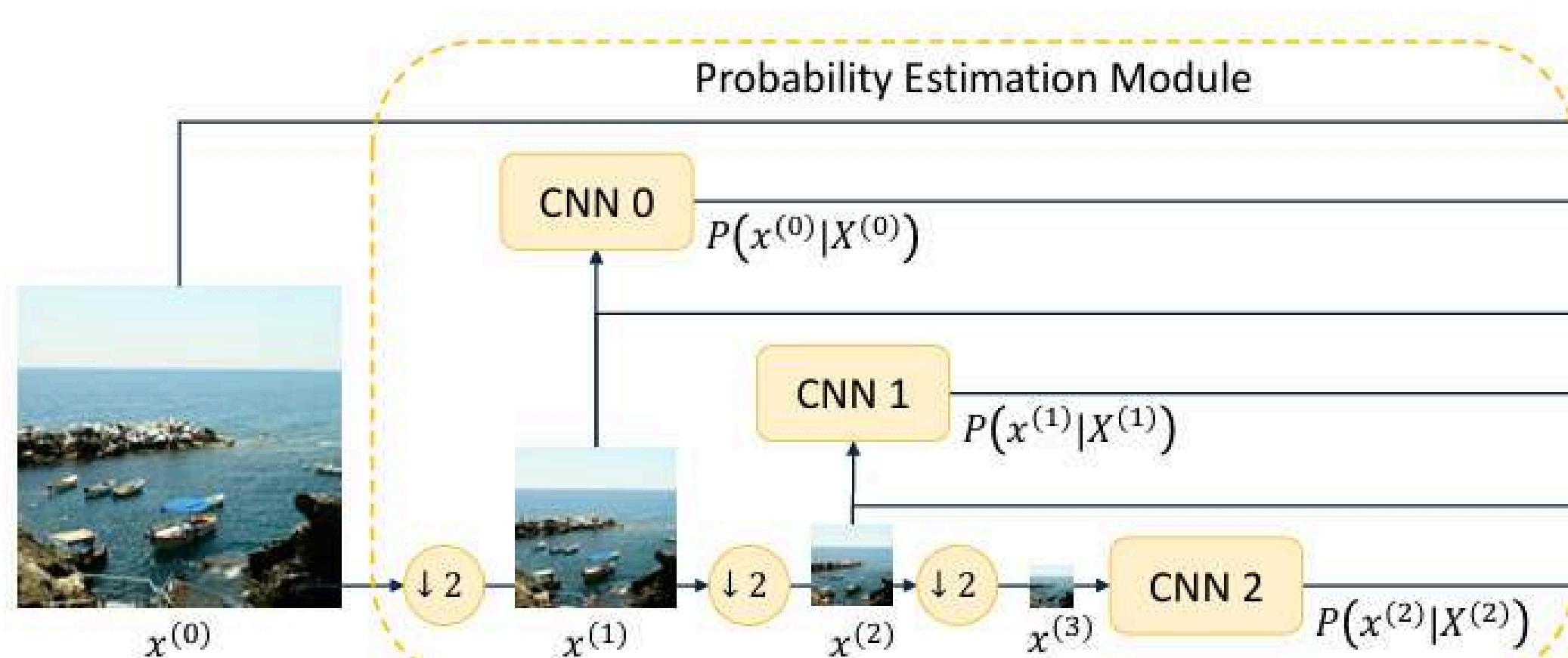
4.1 สร้างภาพหลายความละเอียด (Multi-resolution): เริ่มจากนำภาพที่ต้องการทดสอบ มาอยู่ขนาดลง 3 ระดับด้วยวิธี average pooling ทำให้มีภาพทั้งหมด 4 ความละเอียด ตั้งแต่เล็กสุดไปจนถึงขนาดเดิม



Method

4. กระบวนการทำงาน (ZED: Zero-shot Entropy-based Detector)

4.2 CNN ทำการเรียนรู้ความสัมพันธ์ของพิกเซลในภาพปลอมที่ทดสอบ โดยเข้าถึงโครงสร้างของภาพและการแจกแจงความน่าจะเป็นของพิกเซลในแต่ละขนาดของภาพ เพื่อใช้คำนวณค่าพิกเซลถัดไปได้แม่นยำว่าโอกาสที่พิกเซลบริเวณนี้จะมีการแจกแจงของพิกเซลเป็นอย่างไรจากบริบทรอบๆ

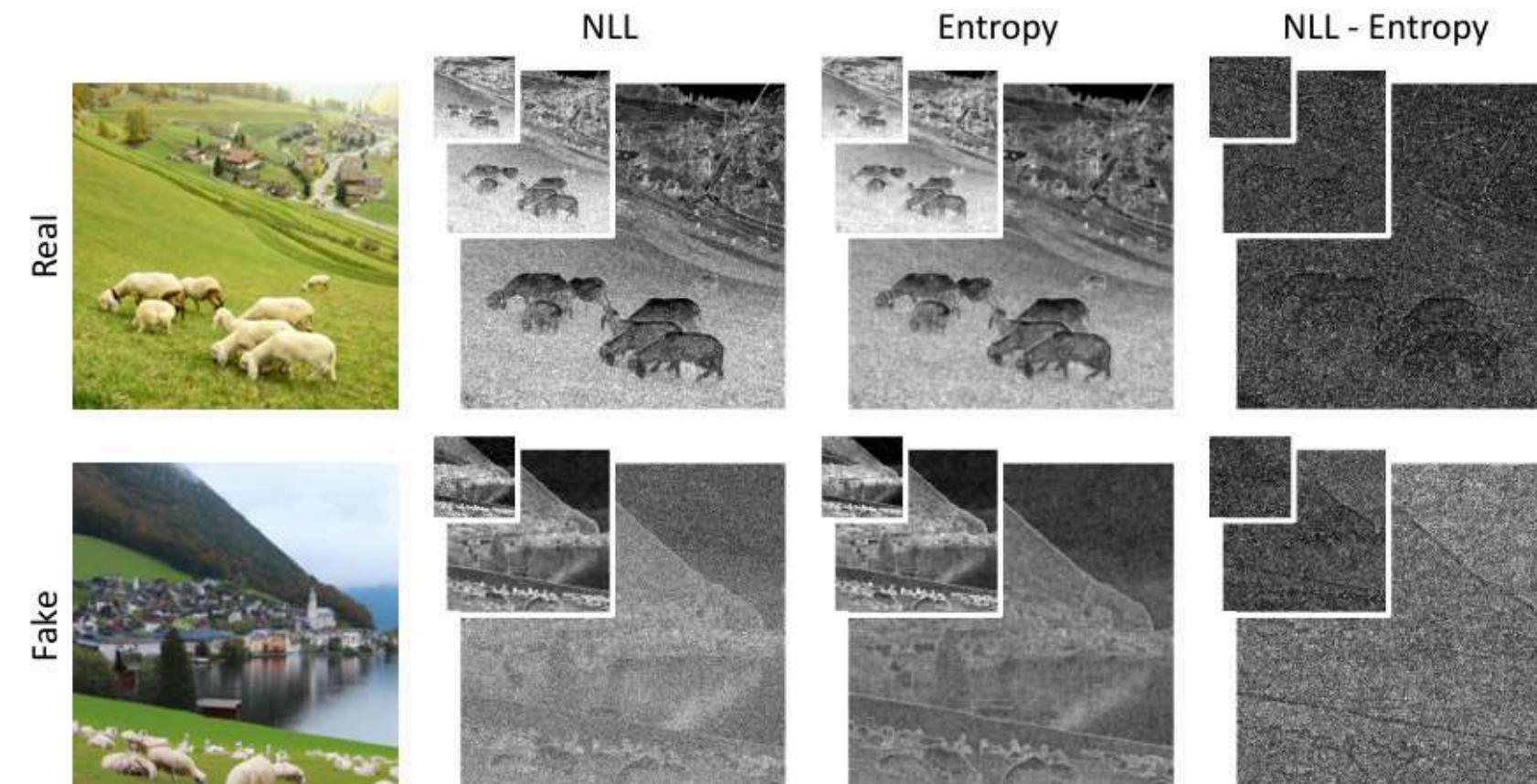


Method

4. กระบวนการทำงาน (ZED: Zero-shot Entropy-based Detector)

4.3 ทำนายและวัดผล: ที่แต่ละระดับความละเอียด (ยกเว้นระดับเล็กสุด) โดยเดล SReC จะใช้ภาพที่ความละเอียดต่ำกว่าเป็น "บริบท" ในการทำนายค่า pixel ของภาพที่ความละเอียดสูงกว่า จากนั้นจะคำนวณค่าทางสถิติอีกมา 2 ตัวคือ:

- **NLL (Negative Log Likelihood):** คือ "ต้นทุนในการเข้ารหัสตามจริง" (actual coding cost) ถ้าค่า NLL สูง แปลว่า pixel นั้นทำนายได้ยาก ไม่เป็นไปตามที่คาด
- **Entropy (H):** คือ "ต้นทุนในการเข้ารหัสที่คาดหวัง" (expected coding cost) หรือพูดง่ายๆ คือความไม่แน่นอนของ pixel นั้นๆ



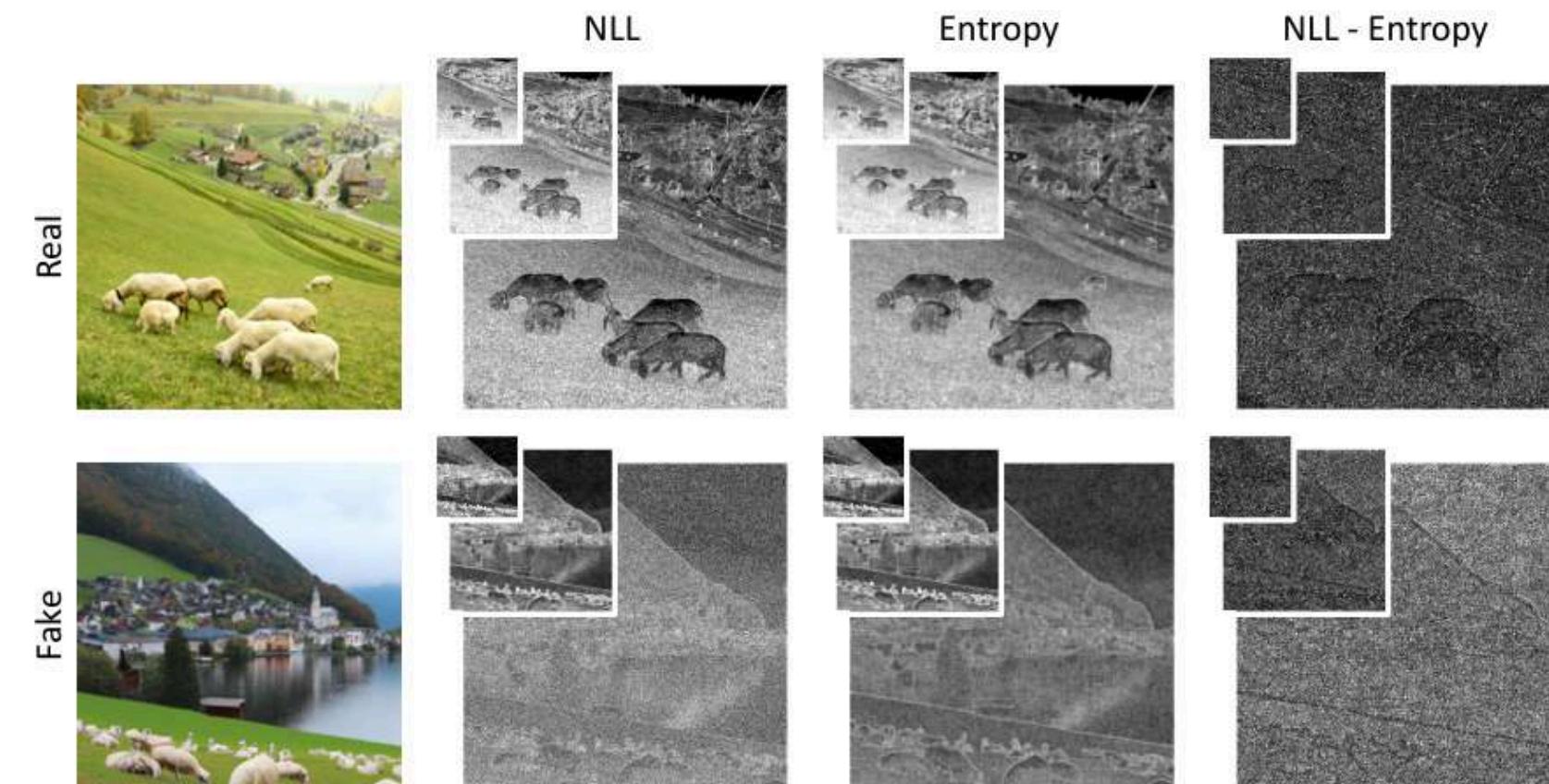
Method

4. กระบวนการทำงาน (ZED: Zero-shot Entropy-based Detector)

4.4 หา "ช่องว่าง" (Coding Cost Gap): การนำค่าหั้งสองมาลบกัน $D = NLL - H$

ถ้าเป็นภาพจริง: โมเดล SReC ที่เรียนรู้จากภาพจริงมา จะทำนายได้ค่อนข้างแม่นยำ ทำให้ค่า NLL กับ H ใกล้เคียงกันมาก ช่องว่าง (gap) นี้ก็จะน้อย

ถ้าเป็นภาพปลอม: ภาพปลอมมีค่าสถิติที่แตกต่างจากภาพจริง ทำให้โมเดล "ประหลาดใจ" และทำนายพลาด ส่งผลให้ค่า NLL สูงกว่า H มาก ช่องว่างนี้ก็จะกว้างขึ้นอย่างเห็นได้ชัด



ตัวชี้วัด

ในขั้นตอนการวิเคราะห์ ZED ไม่ได้ดูแค่ค่า D ค่าเดียว แต่ใช้ ทั้งหมด 3 ตัวชี้วัดหลัก (features) เพื่อทดสอบว่าตัวไหนแยกภาพจริง/ภาพปลอมได้ดีที่สุด เพื่อวิเคราะห์ว่าตัวชี้วัดใดมีผลมากที่สุดต่อ ความสามารถของโมเดลในการแยกแยกภาพจริงกับภาพที่สร้างโดย AI ซึ่งมีตัวชี้วัดดังต่อไปนี้

- $D^{(0)}$ (the 0-level coding cost gap),
- its slope $\Delta^{01} = D^{(0)} - D^{(1)}$,
- and their absolute values.

Method

1. $D^{(0)} = \text{O-level coding cost gap}$

คือ ความแตกต่างของต้นทุนการเข้ารหัส ระหว่างภาพจริงกับภาพปлом ที่ระดับ 0 (ภาพที่ไม่มีการบีบอัด) เช่น ถ้าโมเดลใช้ NLL กับภาพจริงได้ค่าเท่ากับ 2.0 แต่กับภาพปломได้ 3.5 $\rightarrow D^{(0)} = 3.5 - 2.0 = 1.5$ ถ้า $D^{(0)}$ สูง \rightarrow แสดงว่าโมเดล “รับรู้” ความแตกต่างระหว่างภาพจริงกับปломได้ชัด ถ้า $D^{(0)}$ ใกล้ 0 \rightarrow ภาพจริงกับปломแยกกันไม่ออกร

Method

$$2. \Delta^{01} = D^{(0)} - D^{(1)}$$

คือ ความชันของช่องว่างต้นทุนการเข้ารหัส ระหว่างระดับ 0 กับระดับ 1 หมายถึง ความต่างระหว่าง
ภาพจริง-ปلوم ลดลงหรือเพิ่มขึ้น เมื่อความละเอียดลดลง

ตัวอย่าง:

- $D^{(0)} = 1.5$
- $D^{(1)} = 0.5$
- $\Delta^{01} = 1.5 - 0.5 = 1.0 \rightarrow$ แสดงว่าระดับ 0 แยกແຍະได้ดีกว่าชัดเจน

Method

3. Absolute values

การนำเอาค่าที่ได้จาก $D^{(0)}$ หรือ Δ^{01} มาทำให้เป็นบวกหมด เช่น

- ถ้า $D^{(0)} = -2.3 \rightarrow \text{Absolute} = 2.3$
- ใช้เพื่อฉุว่า "มาน้อยแค่ไหน" โดยไม่สนใจว่าบวกหรือลบ

Method

4. กระบวนการทำงาน (ZED: Zero-shot Entropy-based Detector)

4.4 ตัดสินผล: ใช้ค่า "ช่องว่าง" ที่คำนวณได้ โดยเฉพาะที่ความละเอียดสูงสุด ($D(0)$) และผลต่างของช่องว่างระหว่าง 2 ความละเอียด (ΔD_1) มาเป็นตัวตัดสินว่าภาพนั้นเป็นภาพจริงหรือภาพปลอม

4. Result

4.1

Datasets and Metrics

4.2

Ablation Study

4.3

SoTA Comparison

4.4

Limitations

4.1 Datasets and Metrics

- **Datasets**

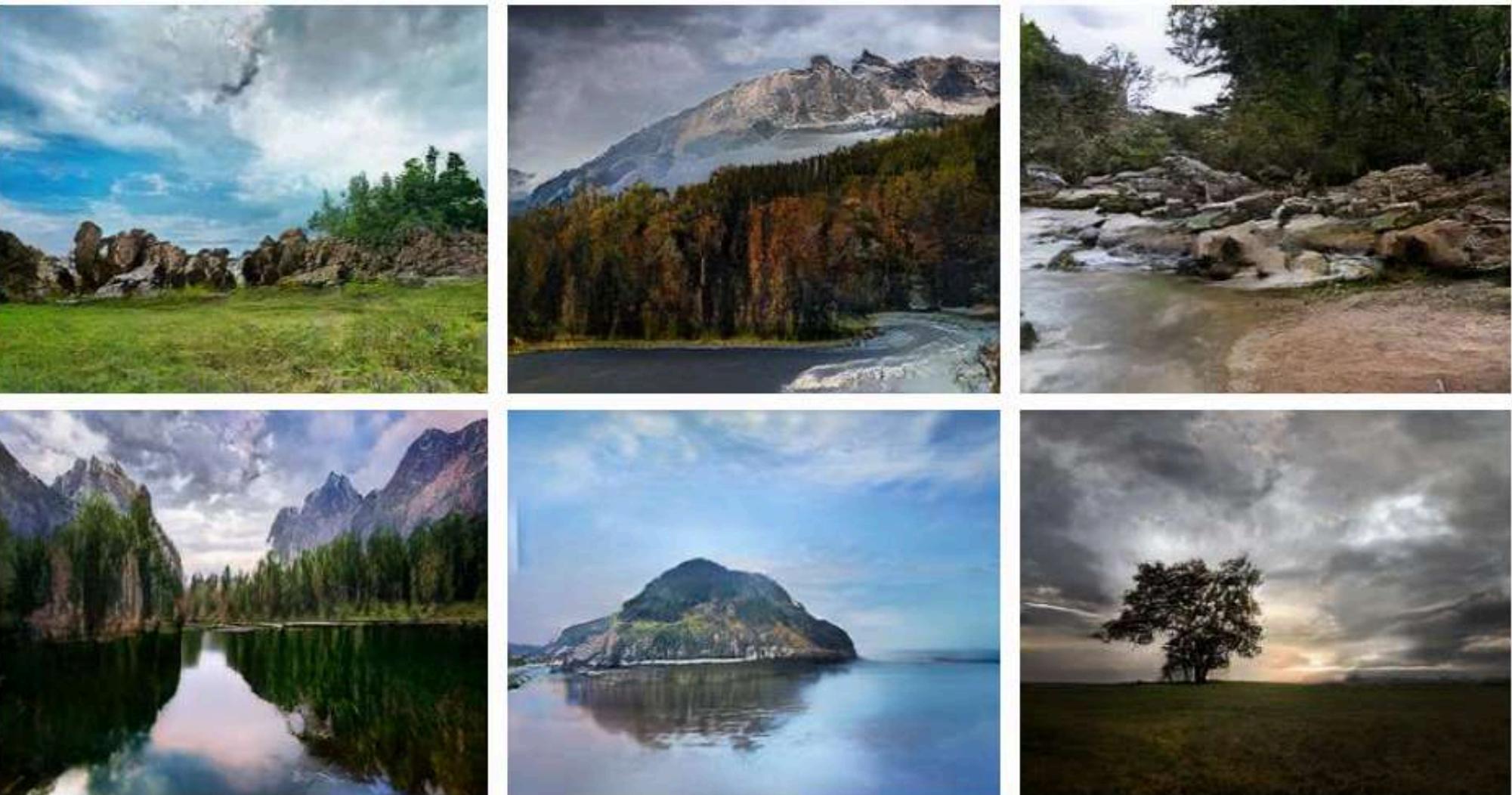
The **29,000 synthetic images** were generated using 19 different generative models, including



4.1 Datasets and Metrics

GauGAN

มีความพิเศษตรงที่สามารถเปลี่ยนภาพวดง่ายๆ ของเราให้กลายเป็นภาพทิวทัศน์ที่ดูสมจริงได้อย่างน่าอัศจรรย์ เช่น ห้องฟ้า ภูเขา หรือแม่น้ำ AI จะให้รายละเอียดที่ซับซ้อนและสวยงามโดยอัตโนมัติ



https://github.com/NVlabs/SPADE/tree/master/datasets/coco_stuff/train_img

4.1 Datasets and Metrics

BigGAN

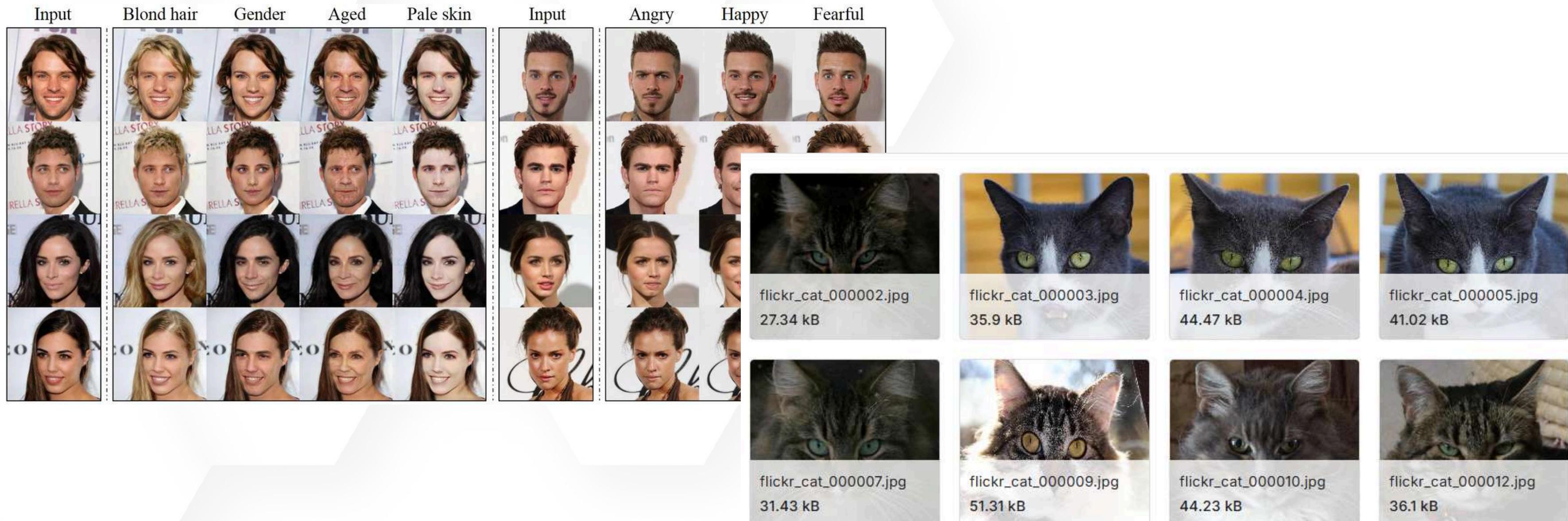
เน้นการสร้างภาพถ่ายที่ดูสมจริงและมีความละเอียดสูงมากๆ โดยจะสร้างภาพตามหมวดหมู่ที่มีอยู่จริง เช่น ภาพสัตว์, อาหาร, หรือสิ่งของต่างๆ



4.1 Datasets and Metrics

StarGAN

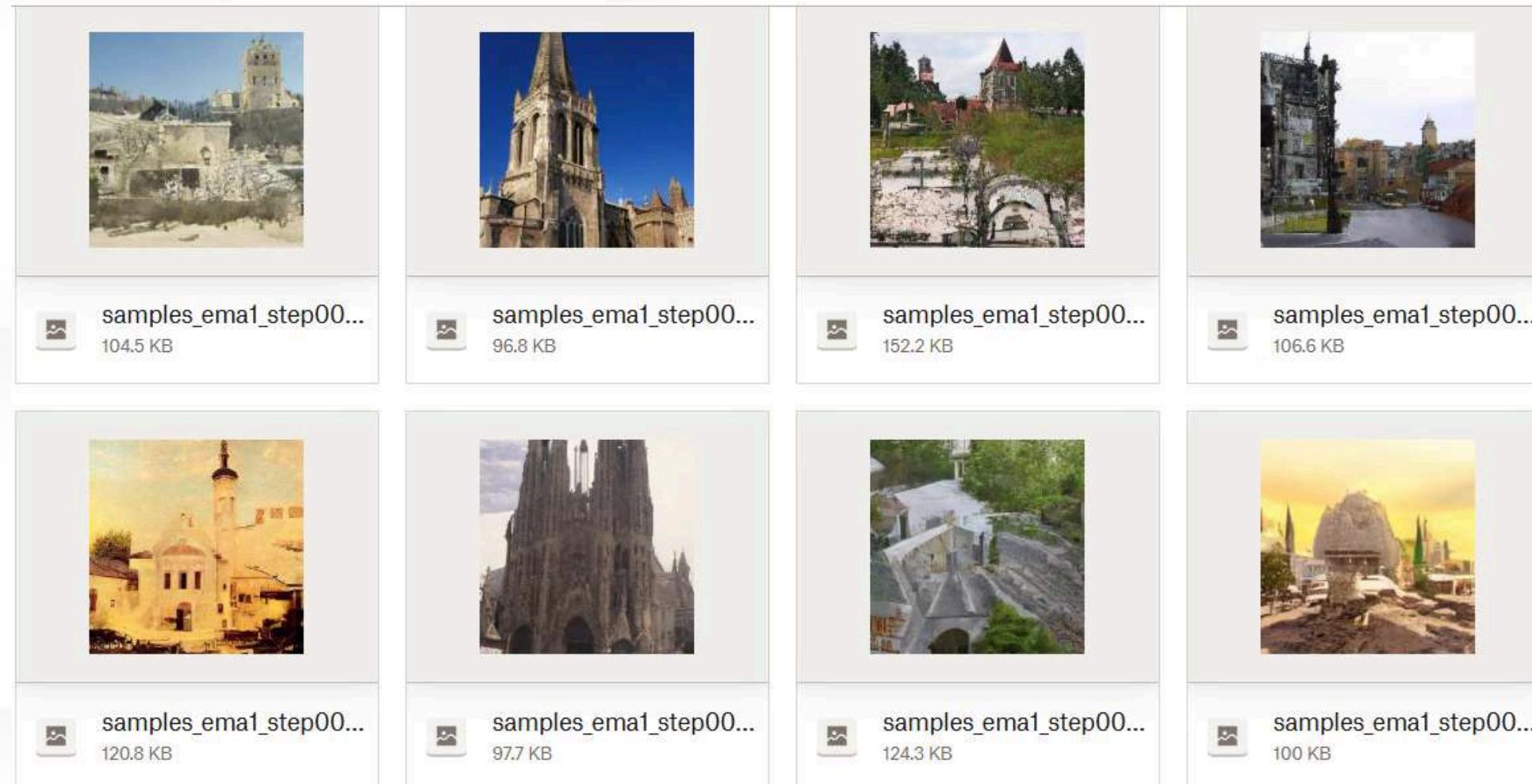
เป็น AI ที่เก่งในเรื่อง "การแปลงโฉมใบหน้า" ตามคำสั่งที่เราต้องการ โดยที่ยังคงเอกลักษณ์ของใบหน้าเดิมไว้ให้มากที่สุดความสามารถหลักของ StarGAN คืออะไร?



4.1 Datasets and Metrics

DDPM

DDPM ไม่ได้เจาะจงว่าจะต้องสร้างรูปประเภทใดประเภทหนึ่ง แต่เน้นที่คุณภาพและความสมจริงขั้นสูงของภาพที่สร้างขึ้นมา ไม่ว่าจะเป็น ภาพคน, สัตว์, สิ่งของ, หรือทิวทัศน์ ถ้าโมเดล DDPM ได้เรียนรู้จากชุดข้อมูลประเภทไหน มันก็จะสามารถสร้างภาพประเภทนั้นๆ ออกมาได้ดีเยี่ยวกัน



4.1 Datasets and Metrics

Stable Diffusion XL

ถูกออกแบบมาเพื่อสร้างภาพที่ดูเหมือนภาพถ่ายจริงมาก ๆ มีความคมชัดสูง จัดการกับแสงเงาได้ดี และเก็บรายละเอียดที่ซับซ้อน เช่น มือ, ใบหน้า, และพื้นผิวต่างๆ ได้ดี



4.1 Datasets and Metrics

Adobe Firefly

The **29,000 synthetic images** were generated using 19 different generative models, including



4.1 Datasets and Metrics

- **Datasets**

The **6,000 real images** used to evaluate the model were collected from high-quality image datasets, including the following sources:

LSUN

FFHQ

ImageNet

COCO

LAION

RAISE

4.1 Datasets and Metrics

- **Datasets**

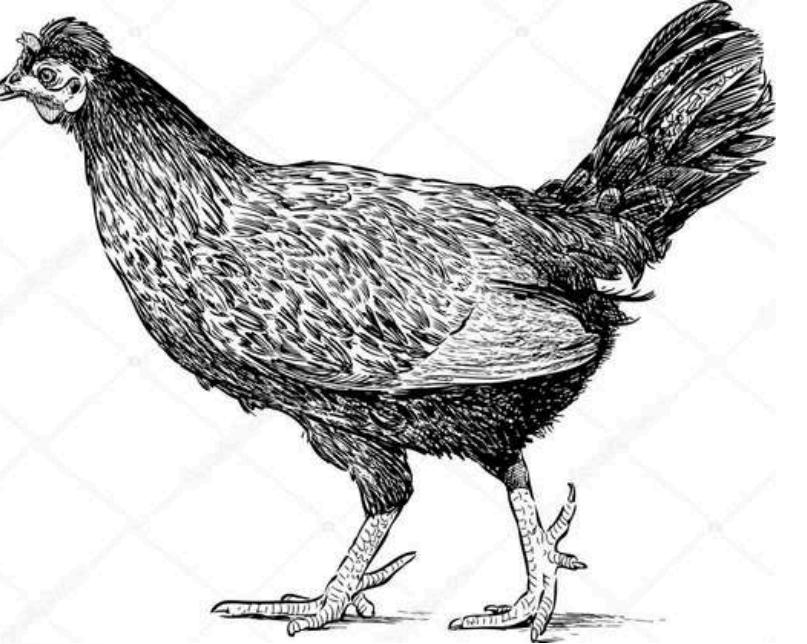
FFHQ



4.1 Datasets and Metrics

- **Datasets**

ImageNet



4.1 Datasets and Metrics

- **Datasets**

COCO

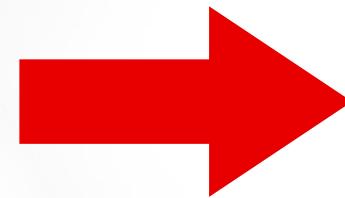


4.1 Datasets and Metrics

Real images are carefully matched in semantic content and compression quality to synthetic ones

- **Datasets**

real images



AI-Generate images

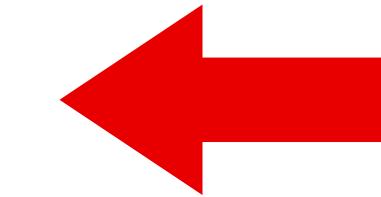
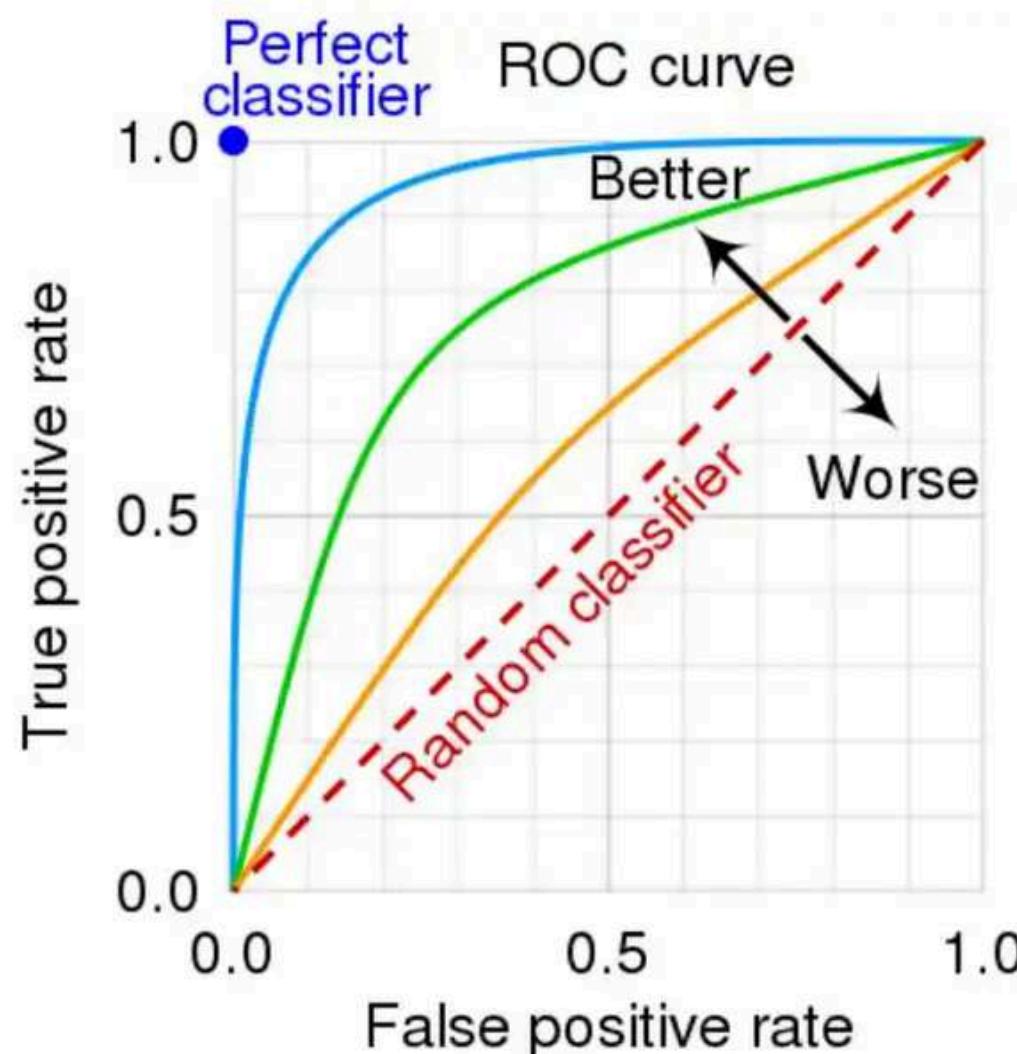


Fig. 4: Examples of real and AI-generated images of different categories used in our experiments. Top: real images from LSUN, FFHQ, ImageNET and COCO. Bottom: generated images from DiffusionGAN, StyleGAN2, DiT and SDXL.

4.1 Datasets and Metrics

- **Evaluation Metrics**

- **AUC (Area Under ROC Curve):** Measures the model's ability to distinguish between real and AI-generated images. The closer to 100%, the better the model's performance. Commonly used in binary classification task



4.1 Datasets and Metrics

• Evaluation Metrics

- **Balanced Accuracy (bACC):** Used when the number of real and fake images is imbalanced. Calculated as the average of true positive rates for each class.
→ Ensures that both real and fake image detection are fairly evaluated.

4.1 Datasets and Metrics

• Evaluation Metrics

- **Threshold Analysis:** The study also analyzes how different decision thresholds affect performance. Threshold = the value used to decide whether an image is “real” or “fake”. This helps assess the model’s robustness and stability in real-world scenarios

4.2 Ablation Study



Fig. 5: Decision statistics. NLL and entropy by themselves are not discriminant (left). Their difference (center) is much more useful for detection, but only at high resolution, $D^{(0)}$, while $D^{(1)}$ is less discriminant and $D^{(2)}$ basically useless. Right box shows histograms of $D^{(0)}$ for real and synthetic images. Note that for GLIDE, $D^{(0)}$ is negative, on the average. Good discrimination is still possible based on the absolute value.

4.2 Ablation Study

Influence of the real class

เพื่อศึกษาว่าการเลือกชุดข้อมูลภาพจริง (real images) ที่ใช้ฝึกโมเดล lossless encoder มีผลต่อประสิทธิภาพของการตรวจจับภาพปลอมจาก AI หรือไม่

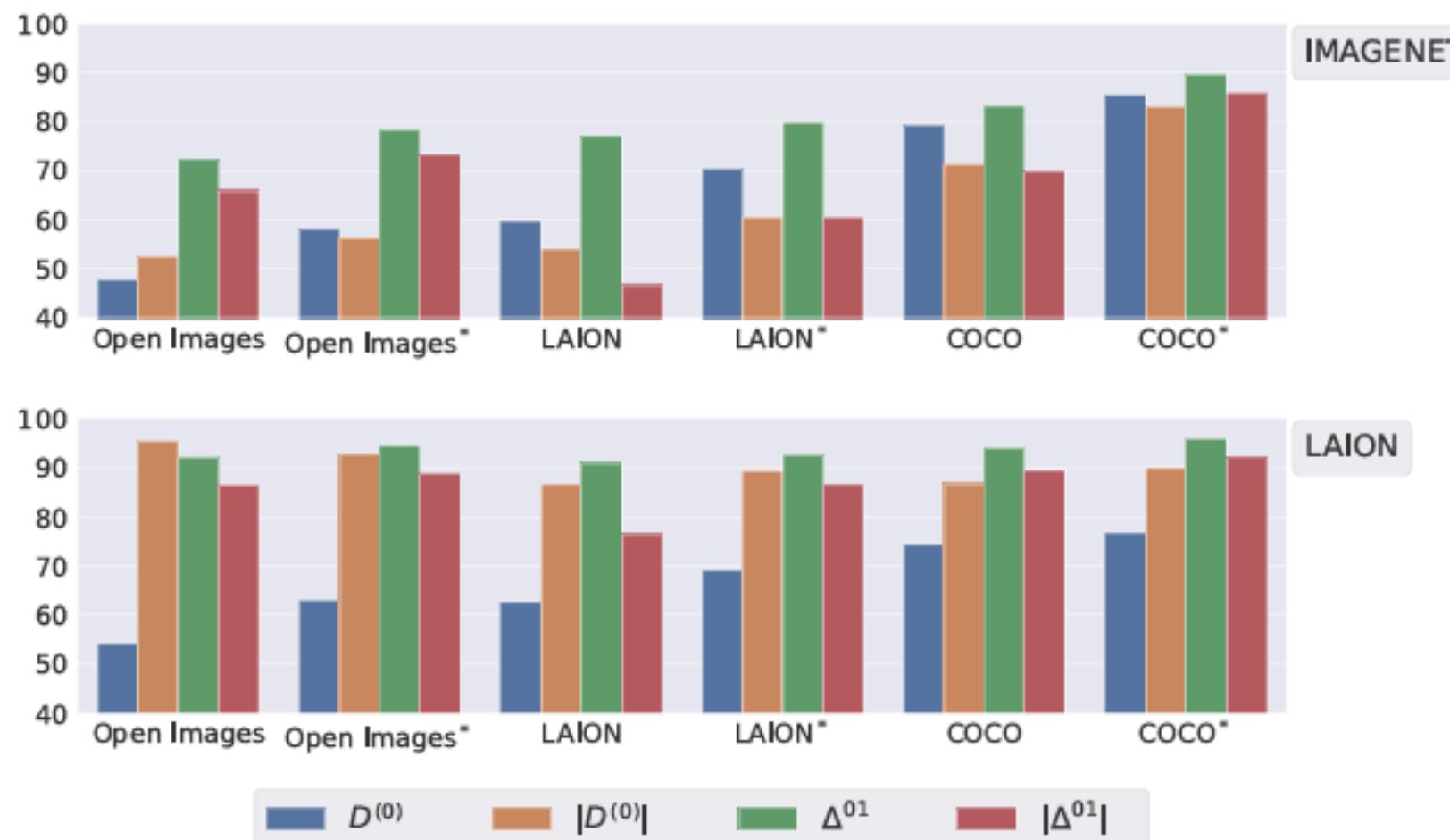


Fig. 6: AUC of proposed method as a function of decision statistic (see Section 3.4) and dataset of real images used to train the lossless encoder: Open Images, LAION, COCO, and their augmented versions (*). Synthetic test images are selected to match the corresponding real test images: ImageNet (top), and LAION (bottom).

4.2 Ablation Study

ตารางแสดงรายละเอียดของวิธีการตรวจจับภาพปลองจากการวิจัยก่อนหน้า โดยเปรียบเทียบในด้านแนวคิด วิธีท伦 (ภาพจริง/ภาพปลองที่ใช้), ขนาดข้อมูล, การใช้การเพิ่มข้อมูล (augmentation) และกลยุทธ์ในการทดสอบ ซึ่งจะใช้เป็นพื้นฐานในการเปรียบเทียบกับวิธีใหม่ของผู้วิจัย เพื่อประเมินว่าแนวทางที่นำเสนอสามารถให้ผลลัพธ์ที่ดีกว่าเพียงใดในบริบทที่หลากหลาย

Table 1: Reference methods. For each one we indicate the key idea, the datasets of real and synthetic images used for training with their sizes, whether or not augmentation is used, the test strategy.

Acronym [ref]	Idea/Approach	Training	Real/Fake	Size(K)	Augment.	Test Strategy
Wang2020 [74]	High diversity	LSUN/ProGAN	360/360	✓	global pooling	
PatchFor. [7]	Patch-based	CelebA,FF/various	84/272		resizing	
Liu2022 [43]	Noise-based	LSUN/ProGAN	360/360	✓	global pooling	
Corvi2023 [10]	No-downsampling	COCO,LSUN/Latent	180/180	✓	global pooling	
LGrad [72]	Gradient-based	LSUN/ProGAN	72/72	✓	resizing	
DIRE [75]	Inversion	LSUN-Bed/ADM	40/40		resizing	
DE-FAKE [67]	Prompt-based	LSUN/Stable Diff.	20/20		resizing	
Ojha2023 [51]	CLIP	LSUN/ProGAN	360/360	✓	cropping	
NPR [71]	Residual	LSUN/ProGAN	72/72		resizing	
AEROBLADE [60]	AE rec. error	- / -	- / -		global distance	

real images, where this latter class comes from ImageNet [15] (Fig.6, top) or LAION [66] (Fig.6, bottom). We can observe that the best and more uniform results across the four decision statistics are obtained using COCO*, while training on Open Images guarantees good performance if the real class is LAION, but bad performance if it is ImageNet. Additional results are included in the supplementary material.

4.3 SoTA Comparison

เปรียบเทียบวิธีใหม่กับ SoTA “State of the Art” ที่มีอยู่ในงานวิจัยก่อนหน้าทั้งหมด 10 วิธี วิธีใหม่ของผู้วิจัยให้ผล AUC เฉลี่ยเกิน 80% ทุกกรณี โดยสูงสุดถึง 90% แสดงให้เห็นว่าวิธีที่นำเสนอมานี้มาและเสถียรกว่าวิธีอื่น ๆ โดยเฉพาะเมื่อต้องรับมือกับข้อมูลที่เปลี่ยนไป

Table 2: AUC for reference and proposed methods. Best score in bold with a 0.5% margin. S = LSUN, F = FFHQ, I = ImageNet, C = COCO, L = LAION, R = RAISE.

	Real data	Wang2020	PatchFor.	Liu2022	Corvi2023	LG Grad	DIRE	DEFAKE	Ojha2023	NPR	AEROBLADE	Ours $D^{(0)}$	Ours $ D^{(0)} $	Ours Δ^{01}	Ours $ \Delta^{01} $
GauGAN	C	98.9	80.8	99.7	83.8	81.6	99.9	43.8	100.	89.1	55.1	99.8	99.8	99.9	99.7
BigGAN	I	92.7	85.5	94.7	83.4	77.2	99.8	59.0	99.6	86.8	51.9	92.3	88.6	95.9	92.6
StarGAN	F	94.7	100.	99.9	95.9	73.9	40.4	45.9	99.7	81.5	84.0	100.	100.	100.	100.
StyleGAN2	S	98.1	83.8	99.7	89.1	99.8	58.3	39.1	96.7	100.	30.0	96.6	96.1	96.7	96.5
	F	94.9	85.1	99.9	58.4	82.7	55.5	47.6	91.0	71.3	60.1	43.1	87.7	41.1	88.7
GigaGAN	I	73.7	61.0	97.3	50.5	76.4	99.9	64.3	94.6	82.4	47.5	72.4	68.1	72.4	68.1
	C	79.5	84.0	99.6	90.9	76.7	99.9	87.9	97.6	95.5	80.6	96.5	94.0	96.7	93.4
Diff.GAN	S	89.8	92.6	99.5	96.6	99.5	49.8	44.8	97.4	100.	43.9	99.4	99.4	99.5	99.5
GALIP	C	89.7	98.2	94.3	87.7	56.7	100.	75.6	98.6	90.7	65.0	98.4	96.3	99.7	99.4
DALL-E	L	66.4	71.7	95.0	98.3	95.2	99.8	55.9	97.3	99.5	24.1	99.2	95.8	98.2	95.4
DDPM	F	31.6	98.4	22.8	100.	9.8	23.1	50.5	77.7	92.4	81.7	76.6	25.2	93.8	79.6
	S	67.6	67.6	70.6	80.3	81.1	52.0	37.4	88.2	94.1	53.1	49.5	53.5	69.4	71.0
ADM	I	61.0	81.9	94.4	81.1	72.7	99.5	69.1	85.3	78.5	80.3	87.8	90.5	95.3	92.1
	C	64.8	97.4	96.3	97.2	81.5	99.9	92.4	88.8	95.4	98.0	47.8	88.5	91.1	91.1
GLIDE	R	32.2	95.0	56.6	86.5	50.6	42.9	92.2	72.8	63.3	87.7	23.2	89.4	51.1	65.1
	L	72.6	74.1	90.8	86.9	90.3	100.	60.2	95.3	99.8	68.7	54.5	84.2	93.8	88.5
DiT	I	58.6	83.1	88.0	100.	56.2	99.6	87.4	77.8	78.4	99.8	89.4	84.3	94.9	91.0
Stable D. 1.4	C	68.2	86.1	95.3	100.	54.7	99.9	93.3	97.9	76.5	99.8	48.4	74.8	54.6	71.4
	R	37.9	61.8	73.4	100.	50.0	37.6	88.0	87.7	43.0	96.9	99.4	98.7	97.0	97.2
Stable D. 2	C	56.5	78.6	94.2	100.	62.8	99.3	97.9	82.3	89.3	99.9	83.0	90.3	84.5	89.1
	R	50.2	38.7	34.8	100.	41.4	35.5	80.7	89.5	44.0	97.4	98.5	96.8	95.8	95.9
SDXL	C	83.8	60.8	89.3	100.	89.3	99.5	94.0	80.0	99.3	87.9	99.9	99.9	99.9	99.8
	R	54.3	68.4	31.1	100.	57.2	47.1	84.4	85.1	76.7	69.7	100.	100.	99.1	99.2
Deep.-IF	C	78.0	62.7	72.2	99.9	68.8	98.9	96.9	92.9	91.6	81.9	91.7	82.3	88.4	79.4
DALL-E 2	C	88.5	52.4	98.9	88.2	78.6	99.9	80.6	97.1	90.0	59.3	100.	100.	100.	99.9
	R	64.8	41.9	70.4	69.4	58.6	44.7	70.9	95.2	39.5	32.8	100.	100.	100.	100.
DALL-E 3	C	65.0	47.3	99.5	100.	88.4	99.9	96.2	86.4	97.7	99.7	99.7	99.5	98.3	98.2
	R	10.9	52.7	0.2	60.8	37.9	47.6	92.4	36.4	48.7	48.3	79.1	66.7	78.0	78.1
Midjourney	R	40.2	57.8	40.7	100.	56.3	51.0	78.1	66.2	77.0	99.0	99.7	99.3	98.5	98.5
Adobe Firefly	R	84.8	49.4	11.8	98.0	40.6	57.4	81.4	97.5	32.1	52.8	73.6	41.2	80.8	80.4
AVG		68.3	73.3	77.0	89.4	68.2	74.6	72.9	88.4	80.1	71.2	83.3	86.4	88.8	90.0

4.3 SoTA Comparison

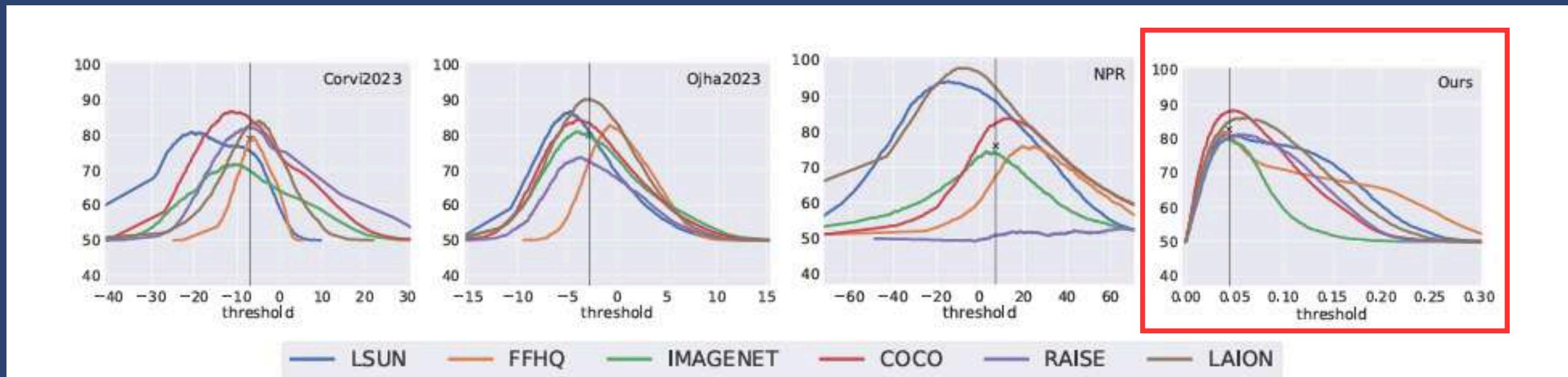


Fig. 7: Balanced accuracy as a function of the detection threshold. For each dataset of real images, we average accuracy over all associated synthetic generators. The dotted vertical line indicates the global optimal threshold and the \times symbol the corresponding accuracy. Note that only for the proposed method all peaks are very close, indicating the presence of a single threshold. Charts for other methods are reported in the Suppl.

กราฟแสดง Balanced Accuracy เมื่อเปลี่ยนค่าการตัดสิน (threshold) ของแต่ละวิธี โดยวิธีที่นำเสนอ มีจุดสูงสุด ของแต่ละเส้นอยู่ใกล้กันมากที่สุด แสดงถึงความเสถียรของวิธีที่ไม่ขึ้นกับชุดข้อมูลจริงที่ใช้แทน สามารถเลือก threshold ค่าเดียวเพื่อใช้กับทุกชุดข้อมูลได้ ตรงข้ามกับวิธีอื่นที่ต้องปรับ threshold ใหม่ในแต่ละกรณีเพื่อให้ได้ผลดีเหมาะสมสำหรับการใช้งานจริง เพราะลดขั้นตอนการปรับแต่งก่อน deployment

THANK YOU