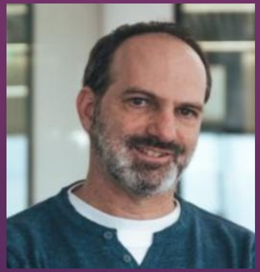
Detecting AI Generated Content using Foundation Models

ROY BETSER

MEIR YOSSEF LEVI

PROF. GUY GILBOA



TECHNION
Israel Institute of Technology

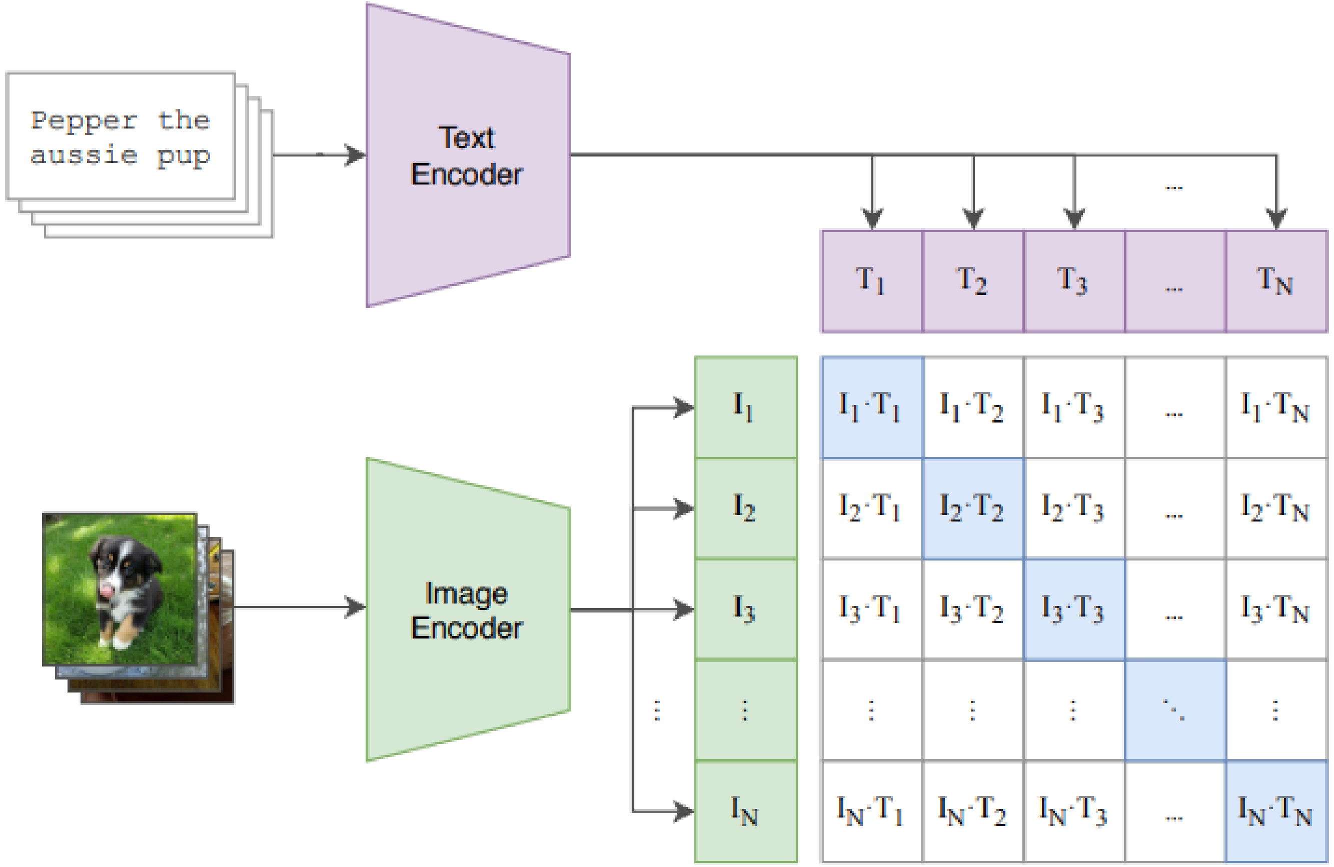
IMVC – 29.06.26

Detecting Generated Content

Real

Generated

What is CLIP?



Classification

Image synthesis

Retrieval

Image Editing

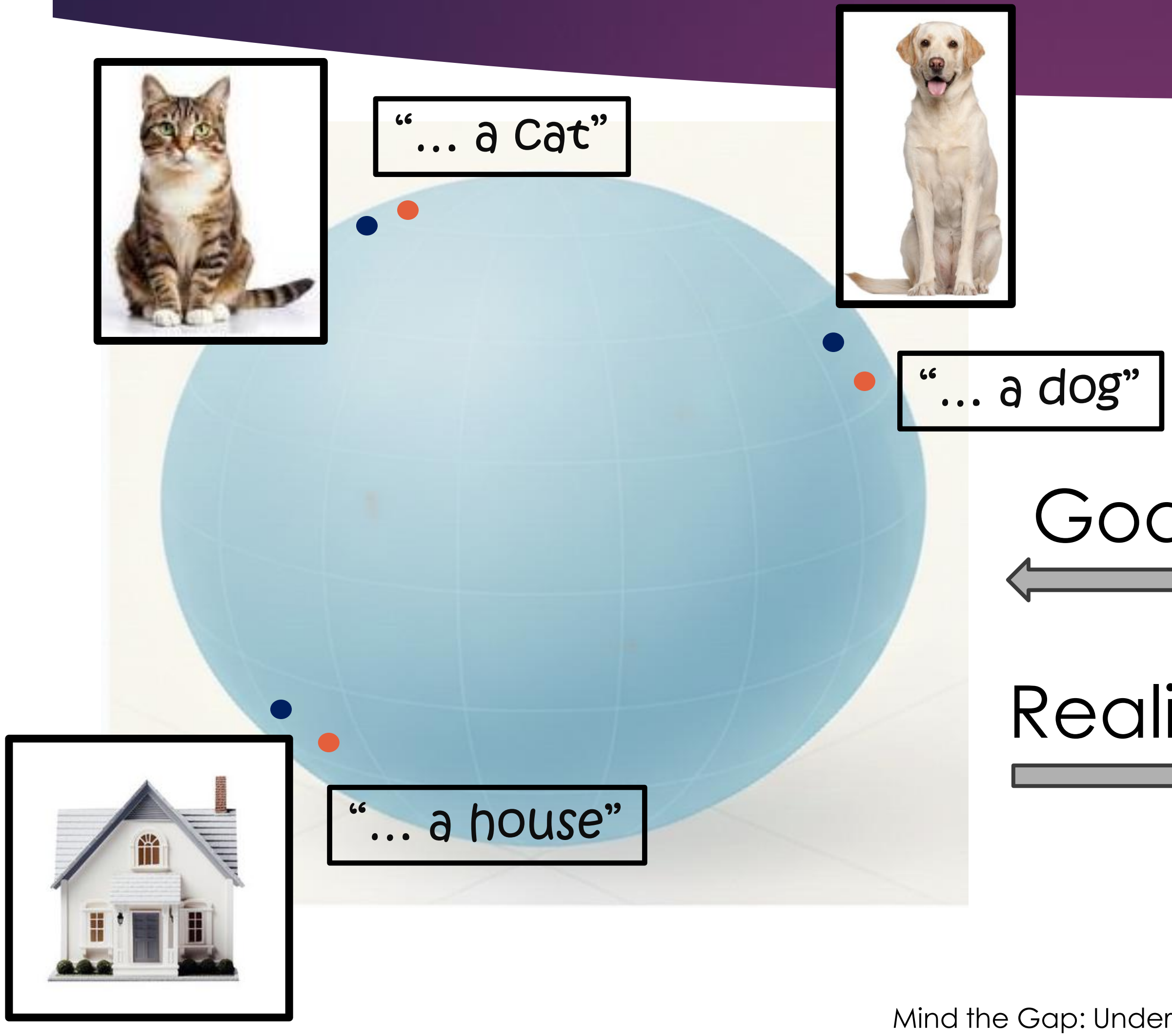
Segmentation

Detection

And many more..

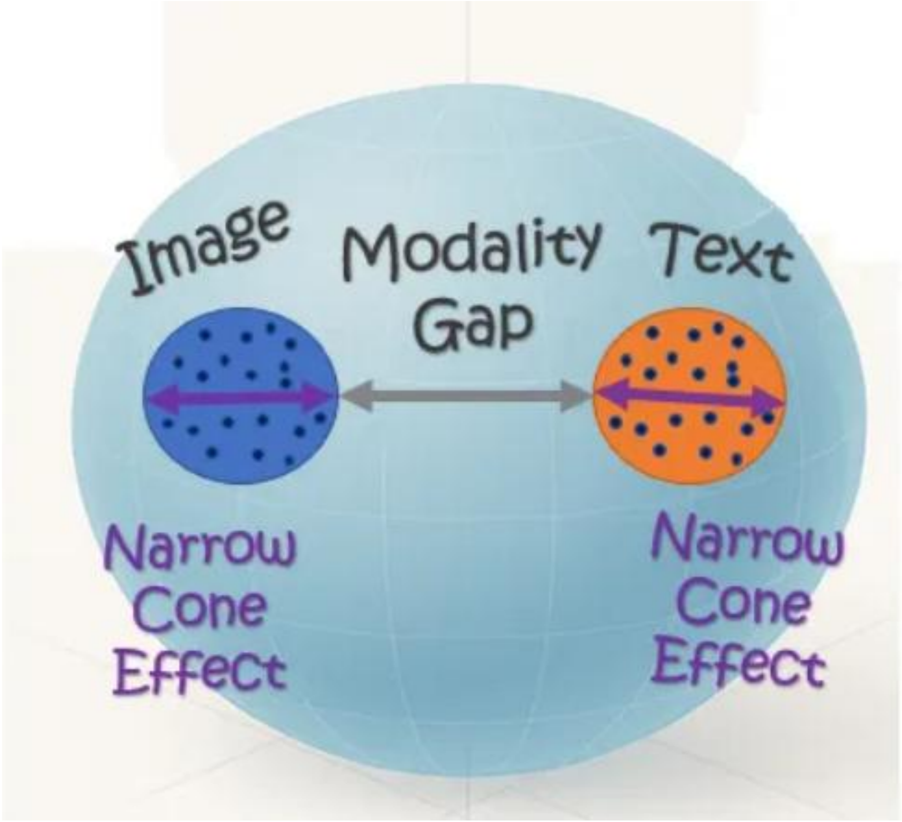
Learning Transferable Visual Models From Natural Language Supervision, Radford et al. 2021

CLIP Latent space



Goal

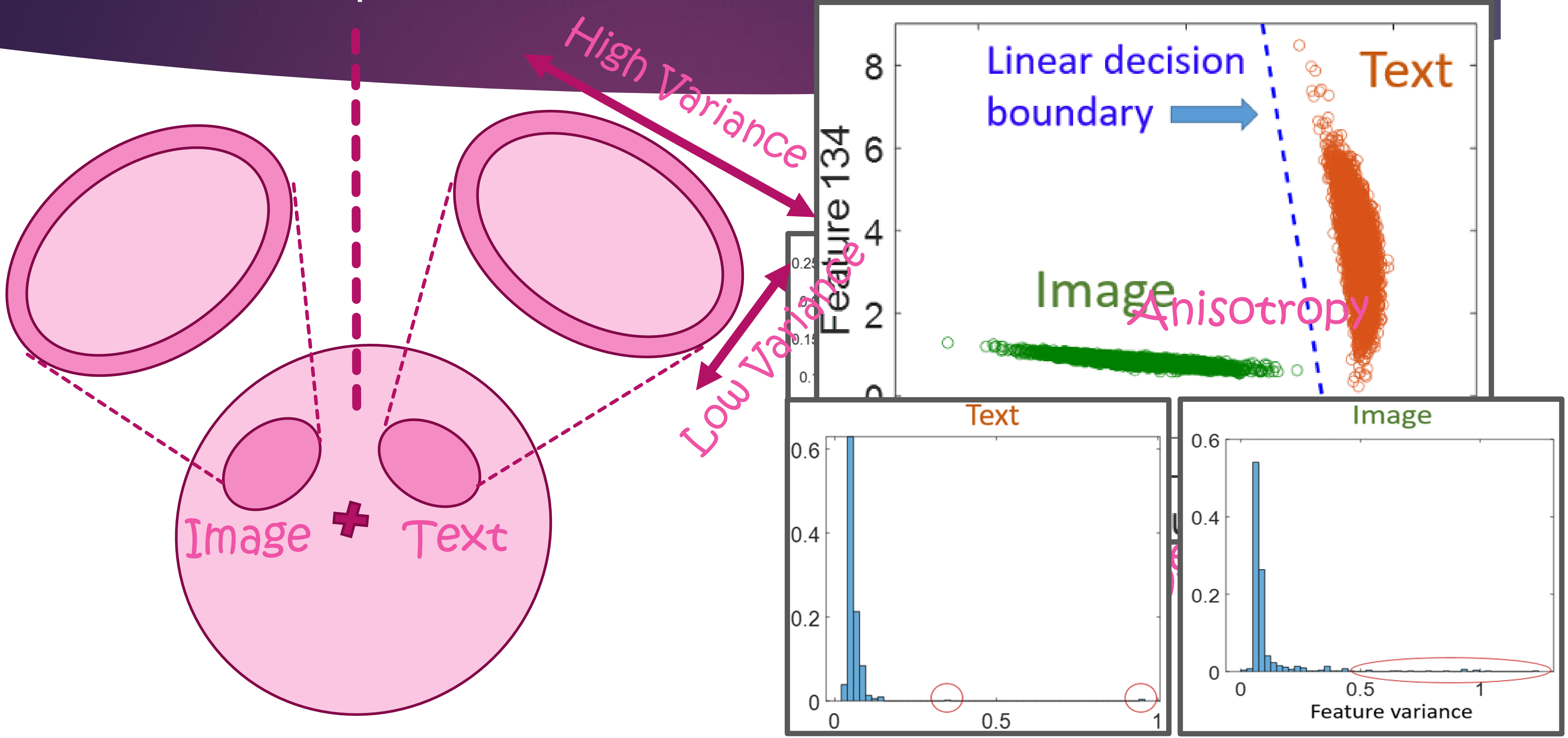
Reality



Mind the Gap: Understanding the Modality Gap in Multi-modal Contrastive Representation Learning, Liang et al. 2022

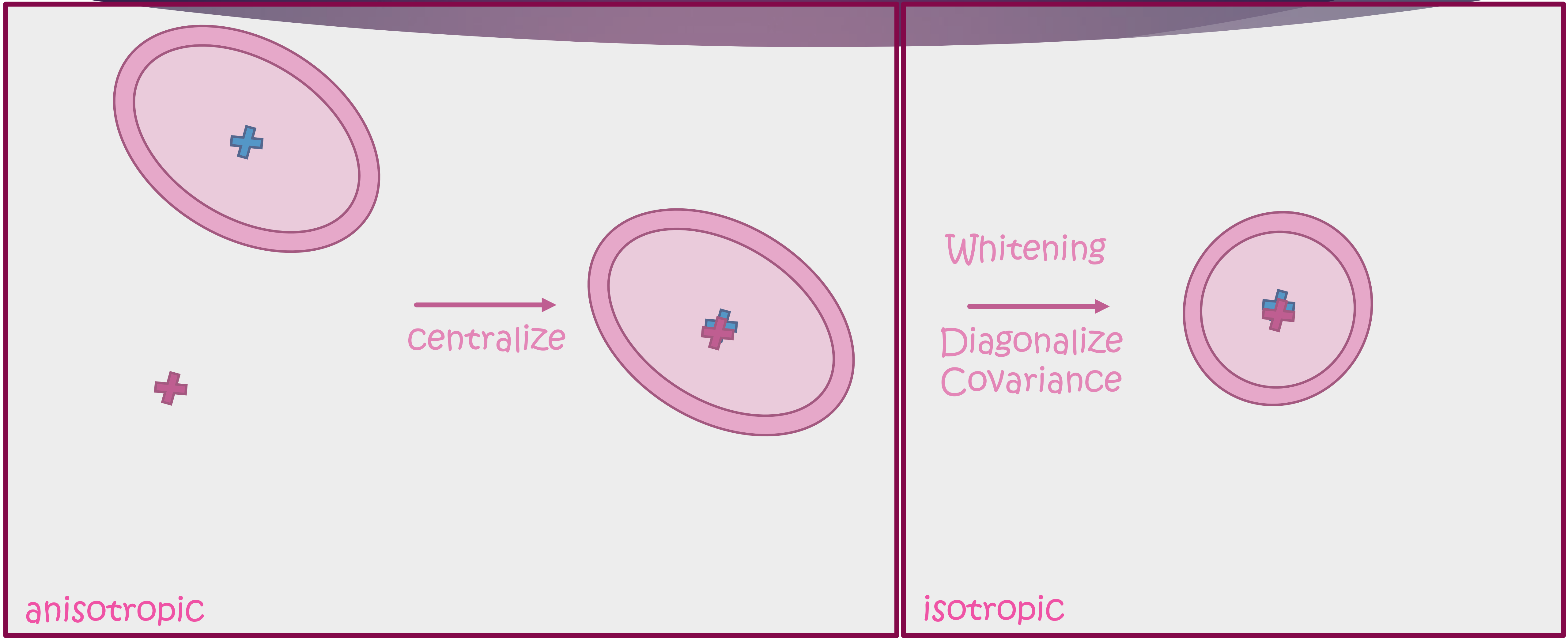
Double Ellipsoid Structure

Scan for
paper details



Whitening CLIP (W-CLIP)

Scan for paper details



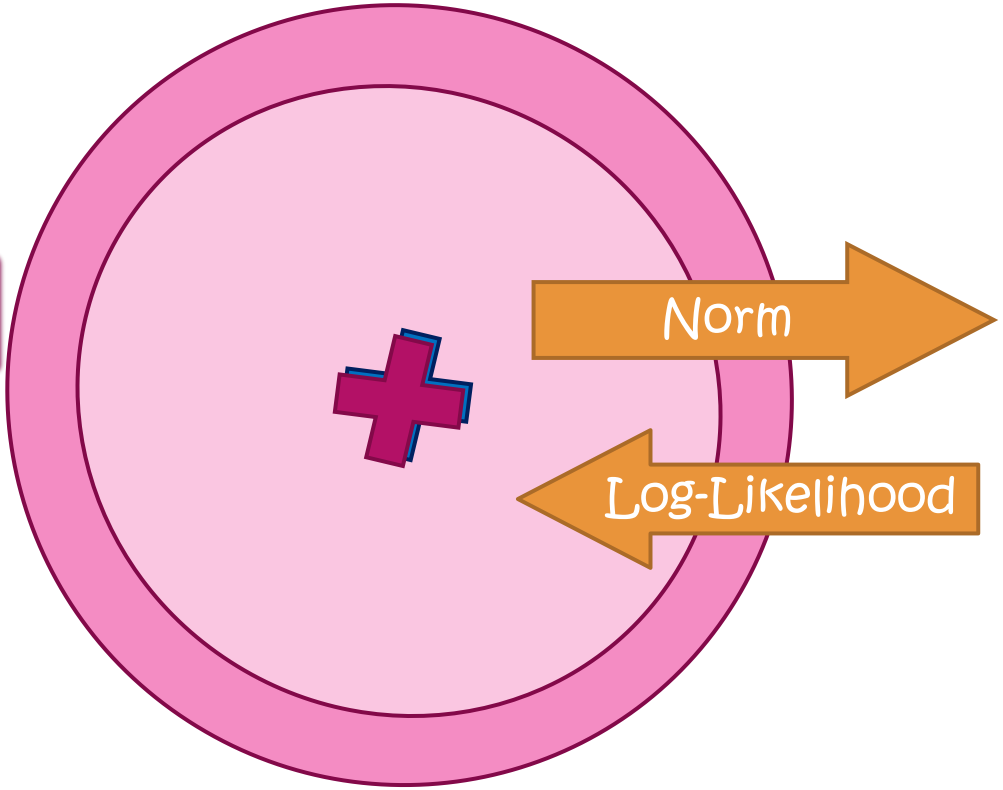
Scan for paper details

Norm as Likelihood

Simple

Fast

Invertible



Approximately Gaussian

	Avg.	Normal Features
	Anderson-Darling (<0.752)	
Image	0.4890	98.3%
Text	0.5926	90.1%
	D'Agostino-Pearson (>0.05)	
Image	0.3624	99.3%
Text	0.2568	99.2%

$$\ell(x) = -\frac{1}{2}(d \cdot log(2\pi) + \|\hat{x}\|^2)$$

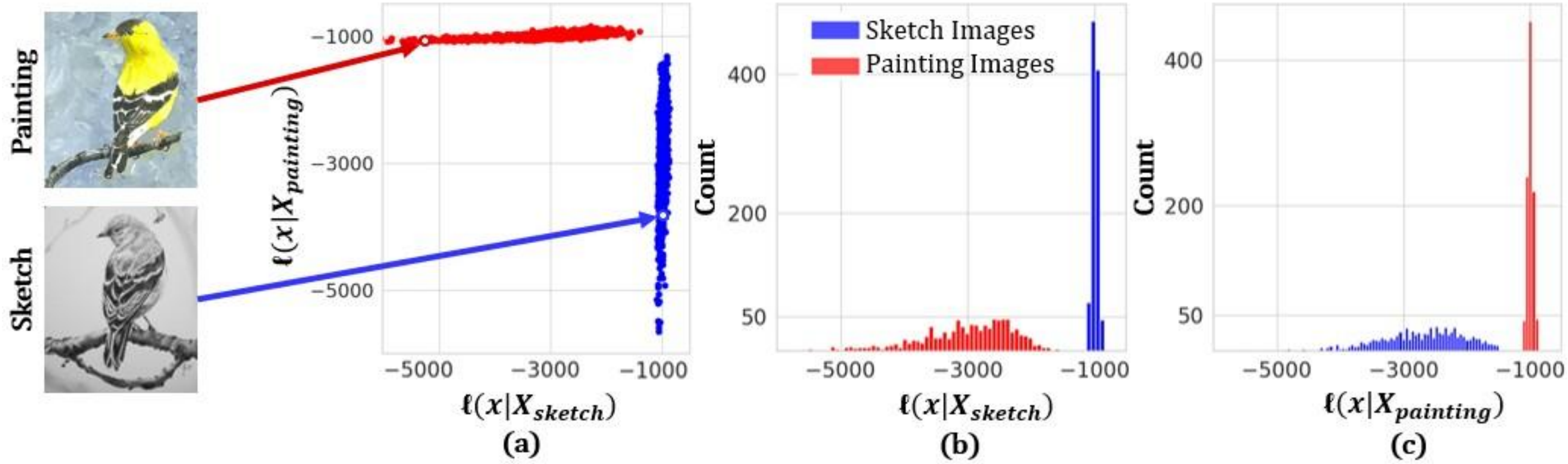
Scan for paper details

FUJITSU

Conditional Likelihood

X - samples from sub-space:

$$\ell(x|X) = -\frac{1}{2}(d \cdot \log(2\pi) + \|W(X,m)\hat{x}\|^2)$$



Scan for paper details

FUJITSU

Detecting Generated Content

Main challenges:

- ❑ Generated content is increasingly photorealistic.
- ❑ New generative models emerge rapidly.

Goal:

Detect generated content by modeling the distribution of real images.

General and Domain-Specific Zero-shot Detection of Generated Images via Conditional Likelihood, WACV 2026

Scan for paper details

FUJITSU

CLIDE – Generated Image Detector

Using only real images for whitening:

Generative Model	AEROBLADE [59]				RIGID [30]				ZED [18]				Manifold Bias [11]				Ours			
	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑
ProGan [34]	0.54	0.49	0.65	0.54	0.45	0.46	0.16	0.48	0.74	0.75	0.7	0.64	**0.96**	**0.97**	**0.88**	**0.88**	0.95	**0.96**	0.87	**0.88**
StyleGan [35]	0.58	0.59	0.63	0.51	0.71	0.72	0.56	0.65	0.75	**0.77**	0.71	0.65	0.68	0.74	0.61	0.68	**0.76**	0.71	**0.75**	**0.73**
StyleGan2 [36]	0.74	0.76	0.68	0.61	0.52	0.54	0.31	0.53	0.75	0.8	0.68	0.61	0.62	0.66	0.48	0.6	**0.8**	**0.82**	**0.73**	**0.68**
BigGAN [9]	0.65	0.61	0.68	0.6	0.62	0.63	0.42	0.57	0.45	0.47	0.63	0.5	0.9	0.9	0.8	0.81	**0.96**	**0.96**	**0.87**	**0.88**
GauGan [51]	0.7	0.71	0.68	0.6	0.32	0.39	0.1	0.45	0.33	0.41	0.6	0.44	**0.98**	**0.98**	**0.91**	**0.9**	0.83	0.85	0.76	0.73
CycleGan [75]	0.76	0.73	0.73	0.69	0.34	0.39	0.05	0.43	0.22	0.38	0.6	0.43	0.97	0.97	0.89	0.9	**0.98**	**0.98**	**0.9**	**0.91**
CRN [15]	0.34	0.51	0.48	0.59	0.25	0.36	0.01	0.42	**0.99**	**0.99**	**0.92**	**0.92**	0.93	0.9	0.88	0.88	0.98	**0.99**	0.91	**0.92**
SD V1.4 [60]	0.47	0.47	0.63	0.51	0.79	0.77	0.63	0.69	0.62	0.65	0.65	0.54	0.71	0.62	0.4	0.57	**0.9**	**0.91**	**0.82**	**0.82**
SD V1.5 [60]	0.49	0.49	0.64	0.52	0.77	0.76	0.63	0.68	0.61	0.63	0.65	0.54	0.72	0.64	0.44	0.58	**0.9**	**0.91**	**0.83**	**0.82**
Guided DM [20]	0.55	0.49	0.66	0.57	0.51	0.51	0.25	0.5	0.58	0.55	0.68	0.59	0.8	0.8	0.66	0.7	**0.87**	**0.87**	**0.79**	**0.78**
LDM 100 [60]	0.48	0.48	0.62	0.48	0.51	0.51	0.25	0.5	0.52	0.56	0.63	0.49	0.84	0.88	0.78	0.79	**0.92**	**0.93**	**0.84**	**0.85**
LDM 200 [60]	0.47	0.48	0.62	0.48	0.51	0.51	0.25	0.5	0.53	0.56	0.63	0.5	0.85	0.88	0.79	0.8	**0.92**	**0.93**	**0.85**	**0.85**
Glide 50 27 [48]	0.06	0.32	0.59	0.42	0.51	0.51	0.25	0.5	**0.98**	**0.98**	**0.91**	**0.92**	0.93	0.94	0.86	0.86	0.91	0.92	0.84	0.84
Glide 100 27 [48]	0.08	0.32	0.6	0.43	0.51	0.51	0.25	0.5	**0.98**	**0.98**	**0.91**	**0.92**	0.92	0.93	0.84	0.84	0.91	0.92	0.83	0.83
Glide 100 10 [48]	0.04	0.31	0.59	0.42	0.51	0.51	0.25	0.5	**0.98**	**0.98**	**0.92**	**0.92**	0.93	0.93	0.85	0.85	0.91	0.92	0.83	0.83
ADM [20]	0.45	0.43	0.65	0.56	0.56	0.57	0.39	0.56	0.47	0.46	0.65	0.55	0.62	0.57	0.32	0.53	**0.89**	**0.88**	**0.84**	**0.84**
DALL·E 3 [6]	0.35	0.41	0.61	0.45	0.51	0.51	0.25	0.5	0.45	0.5	0.62	0.47	0.82	0.84	0.71	0.74	**0.96**	**0.96**	**0.88**	**0.89**
Midjourney [45]	0.29	0.37	0.6	0.43	0.65	0.66	0.48	0.6	**0.92**	**0.93**	0.84	0.84	0.56	0.53	0.27	0.51	0.9	0.84	**0.85**	**0.85**
VDQM [28]	0.51	0.47	0.66	0.57	0.53	0.54	0.32	0.53	0.46	0.46	0.63	0.49	0.83	0.83	0.7	0.73	**0.92**	**0.93**	**0.84**	**0.84**
Wukong [46]	0.48	0.47	0.62	0.49	0.63	0.6	0.37	0.55	0.39	0.45	0.6	0.44	0.73	0.69	0.5	0.62	**0.95**	**0.96**	**0.88**	**0.88**
All	0.52	0.48	0.64	0.53	0.51	0.53	0.28	0.52	0.69	0.66	0.69	0.62	0.85	0.88	0.76	0.78	**0.91**	**0.91**	**0.84**	**0.84**

General and Domain-Specific Zero-shot Detection of Generated Images via Conditional Likelihood, WACV 2026

Scan for paper details

FUJITSU

Damaged Cars








General and Domain-Specific Zero-shot Detection of Generated Images via Conditional Likelihood, WACV 2026

Scan for paper details

Damaged Cars

Real

Fake

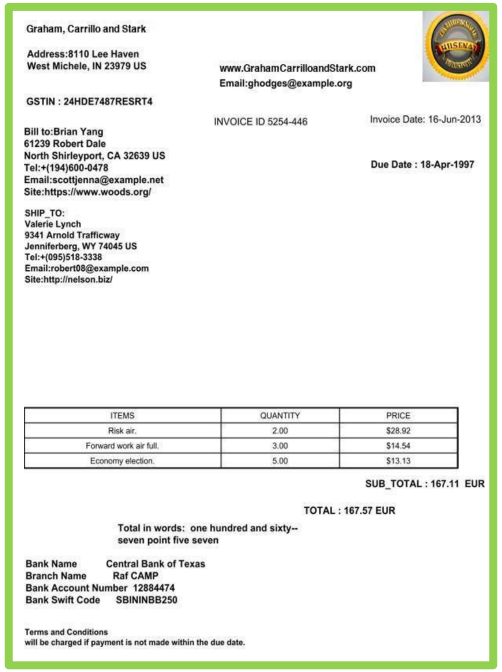
Generative Model	AEROBLADE [59]				RIGID [30]				ZED [18]				Manifold Bias [11]				Ours			
	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑
Kandinsky2.1 [58]	0.29	0.38	0.6	0.43	0.3	0.38	0.06	0.44	0.68	0.66	0.68	0.61	0.6	0.64	0.48	0.61	0.97	0.98	0.93	0.93
SDXL [53]	0.34	0.4	0.61	0.45	0.3	0.38	0.06	0.44	0.83	0.85	0.76	0.73	0.32	0.42	0.15	0.49	0.97	0.97	0.91	0.91
DALL·E 3 [6]	0.39	0.45	0.6	0.44	0.4	0.42	0.11	0.42	0.15	0.33	0.59	0.42	0.14	0.33	0.03	0.44	0.99	0.99	0.93	0.93
Midjourney [45]	0.46	0.47	0.62	0.47	0.51	0.5	0.23	0.5	0.78	0.75	0.74	0.71	0.4	0.44	0.13	0.48	0.93	0.94	0.87	0.86
UnCLIP [57]	0.22	0.35	0.6	0.43	0.69	0.67	0.46	0.6	0.31	0.38	0.61	0.46	0.69	0.68	0.48	0.61	0.8	0.8	0.72	0.67
All	0.33	0.4	0.61	0.44	0.45	0.48	0.21	0.49	0.61	0.54	0.68	0.61	0.47	0.53	0.3	0.53	0.92	0.92	0.85	0.84

General and Domain-Specific Zero-shot Detection of Generated Images via Conditional Likelihood, WACV 2026

Scan for paper details

FUJITSU

Invoice Documents



Graham, Carrillo and Stark

Address:8110 Lee Haven
West Michele, IN 23979 US

www.GrahamCarrilloandStark.com
Email:ghodges@example.org

GSTIN : 24HDE7487RESRT4

INVOICE ID 5254-446

Invoice Date: 16-Jun-2013

Bill to:Brian Yang
61239 Robert Dale
North Shirleyport, CA 32639 US
Tel:+(194)600-0478
Email:scottjenna@example.net
Site:https://www.woods.org/

Due Date : 18-Apr-1997

SHIP_TO:
Valerie Lynch
9341 Arnold Trafficway
Jenniferberg, WY 74045 US
Tel:+(095)518-3338
Email:robert08@example.com
Site:http://nelson.biz/

ITEMS	QUANTITY	PRICE
Risk air.	2.00	$28.92
Forward work air full.	3.00	$14.54
Economy election.	5.00	$13.13

SUB_TOTAL : 167.11 EUR

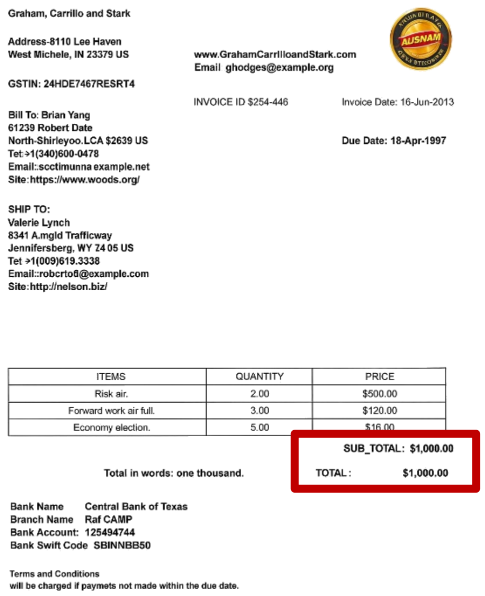
TOTAL : 167.57 EUR

Total in words: one hundred and sixty--
seven point five seven

Bank Name Central Bank of Texas
Branch Name Raf CAMP
Bank Account Number 12884474
Bank Swift Code SBININBB250

Terms and Conditions
will be charged if payment is not made within the due date.

Original Image



Graham, Carrillo and Stark

Address-8110 Lee Haven
West Michele, IN 23379 US

www.GrahamCarrilloandStark.com
Email ghodges@example.org

GSTIN: 24HDE7467RESRT4

INVOICE ID 5254-446

Invoice Date: 16-Jun-2013

Bill To: Brian Yang
61239 Robert Date
North-Shirleyoo.LCA $2639 US
Tet >1(340)600-0478
Email:scctimunna example.net
Site:https://www.woods.org/

Due Date: 18-Apr-1997

SHIP TO:
Valerie Lynch
8341 A.mgld Trafficway
Jennifersberg, WY 74 05 US
Tet >1(009)619.3338
Email::robcrto8@example.com
Site:http://nelson.biz/

ITEMS	QUANTITY	PRICE
Risk air.	2.00	$500.00
Forward work air full.	3.00	$120.00
Economy election.	5.00	$16.00

SUB_TOTAL: $1,000.00
TOTAL: $1,000.00

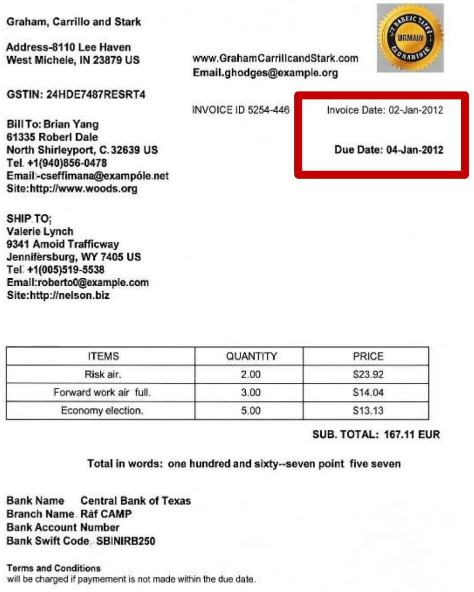
Total in words: one thousand.

Bank Name Central Bank of Texas
Branch Name Raf CAMP
Bank Account: 125494744
Bank Swift Code SBINNBB50

Terms and Conditions
will be charged if payments not made within the due date.

Change price

Change date



Graham, Carrillo and Stark

Address-8110 Lee Haven
West Michele, IN 23879 US

www.GrahamCarrilloandStark.com
Email.ghodges@example.org

GSTIN: 24HDE7487RESRT4

INVOICE ID 5254-446

Invoice Date: 02-Jan-2012

Due Date: 04-Jan-2012

Bill To: Brian Yang
61335 Roberl Dale
North Shirleypoft, C.32639 US
Tel. +1(940)856-0478
Email-cseffimana@exampôle.net
Site:http://www.woods.org

SHIP TO;
Valerie Lynch
9341 Amoid Trafficway
Jennifersburg, WY 7405 US
Tel. +1(005)519-5538
Email:roberto0@example.com
Site:http://nelson.biz

ITEMS	QUANTITY	PRICE
Risk air.	2.00	$23.92
Forward work air full.	3.00	$14.04
Economy election.	5.00	$13.13

SUB. TOTAL: 167.11 EUR

Total in words: one hundred and sixty--seven point five seven

Bank Name Central Bank of Texas
Branch Name Raf CAMP
Bank Account Number
Bank Swift Code SBINIRB250

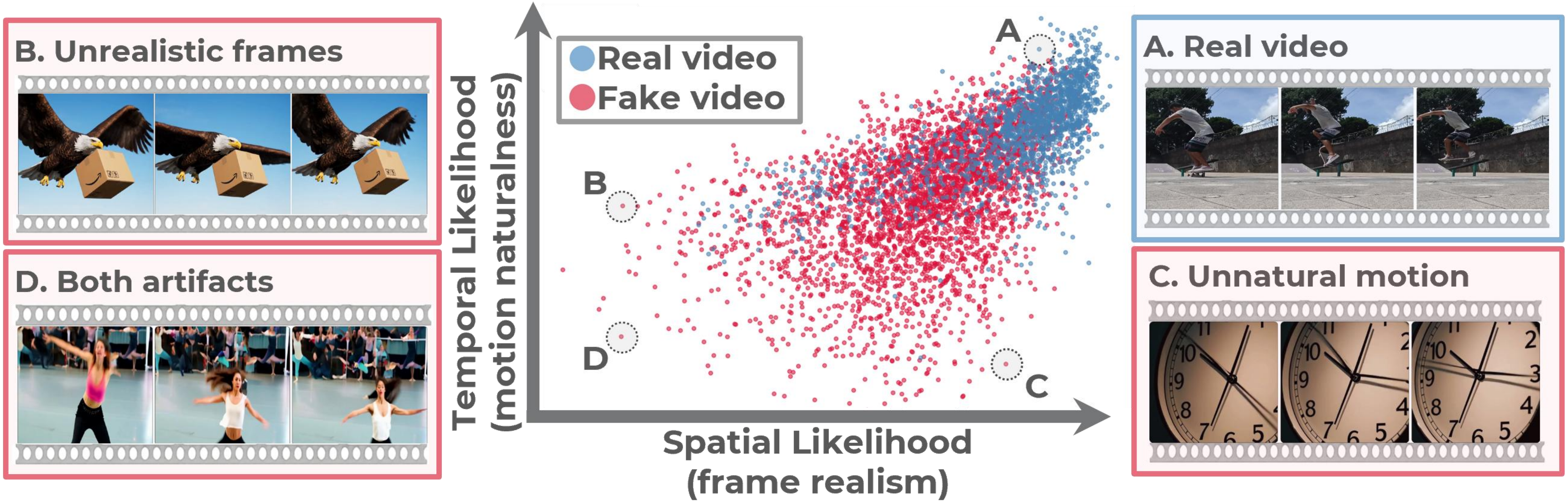
Terms and Conditions
will be charged if paymement is not made within the due date.

Generative Model	AEROBLADE [59]				RIGID [30]				ZED [18]				Manifold Bias [11]				Ours			
	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑	AUC ↑	AP ↑	F1 ↑	Acc ↑
GPT-Image-1 [55]	0.53	0.62	0.56	0.47	0.57	0.67	0.12	0.53	0.6	0.56	0.65	0.54	0.61	0.58	0.55	0.62	**0.91**	**0.92**	**0.82**	**0.82**

Training-free Detection of Generated Videos via Spatial-Temporal Likelihoods, CVPR 2026

Scan for paper details

Generated Video Detection



Summary and takeaways

- CLIP has double ellipsoid structure.
- Vision encoders learn approximately Gaussian representations.
- Understanding latent geometry enables downstream applications.
- Foundation models naturally separate real and generated content.
- Generated content detection can rely only real samples.

Thank you!

Questions?