CHROMA: DETECTING AI-GENERATED IMAGES THROUGH INTER-CHANNEL COLOR-SPACE CORRELATIONS

AUTHORS

JUAN PABLO SOTELO[1]
PABLO MUSÉ[12]
MARINA GARDELLA[2]

1 - IIE, FACULTAD DE INGENIERÍA, UNIVERSIDAD DE LA REPÚBLICA, URUGUAY
2 - UNIVERSITÉ PARIS-SACLAY, ENS PARIS-SACLAY, CNRS, CENTRE BORELLI, FRANCE

UNIVERSIDAD DE LA REPUBLICA URUGUAY
FACULTAD DE INGENIERÍA
école normale supérieure paris—saclay
CENTRE BORELLI

ICPR 2026 · LYON

GENERATORS NOW MATCH THE APPEARANCE OF REAL PHOTOS

- Photorealism is essentially solved, appearance is no longer a reliable cue
- Detectors must rely on low-level statistical traces
- But generally, those traces are generator-specific → detectors are brittle across unseen models

THE CORE ASYMMETRY

Natural Images

Physical measurement chain

Sensor noise, CFA, CC, compression

Traces tied to optics, electronics and processing algorithms

Generated Images

Learned computational pipeline

Architecture, training data, losses

Traces tied to optimization

Forensic strategy: target low-level statistical differences

COLOR CUES ARE UNDEREXPLOITED

The Gap

- Prior work: marginal color statistics: histograms, palettes, chrominance residuals
- Unexplored: the dependence between channels

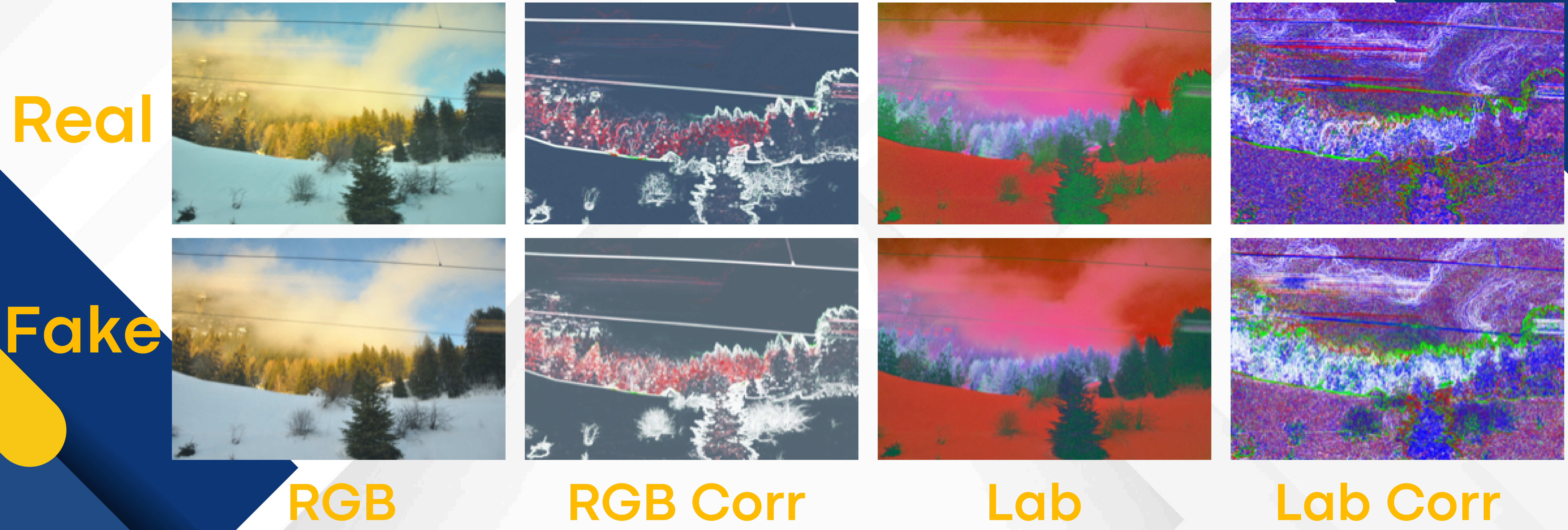
Contributions

- LPIPS doesn't constrain channel dependence uniformly
- Correlations are generator-specific — strongest in Lab
- CHROMA: +correlation maps on a stock ResNet-50 → 85.6 % AUC / 18 generators

LOCAL INTER-CHANNEL CORRELATION

- Sliding **7×7** window, per pixel
- **Pearson** coefficient for each channel pair
- → **3 correlation maps** per color space (e.g. Lab: L–a, a–b, L–b)
- Computed in **RGB, Lab, HSV, YUV**

SAME SCENE, DIFFERENT CORRELATION STRUCTURE



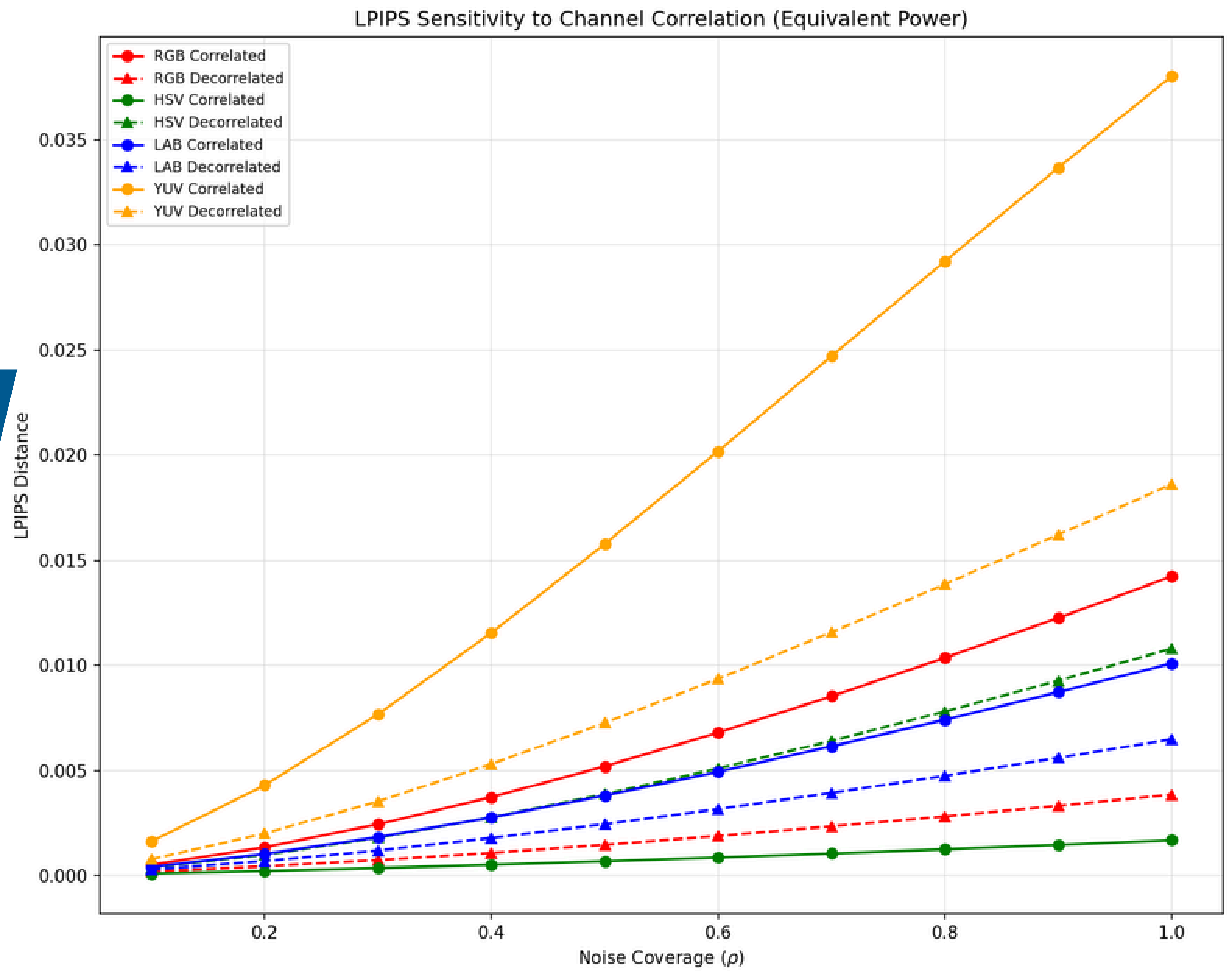
DOES LPIPS PENALIZE CHANGES IN CHANNEL DEPENDENCE?

- Zero-mean Gaussian noise, σ = 0.005, on a fraction ρ of pixels
- **Correlated**: one field n added to all three channels
- **Decorrelated**: independent n_R, n_G, n_B
- Repeated in RGB, Lab, HSV, YUV, with variance rescaled so the effective RGB perturbation power is equal
- Measured on RAISE-1K

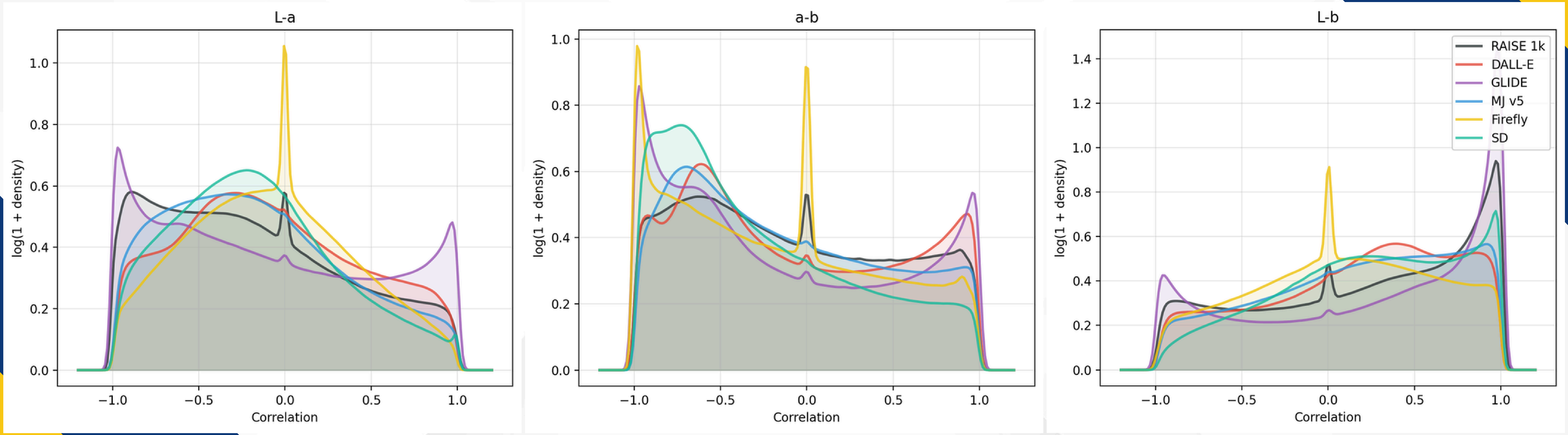
LPIPS TREATS CHANNEL DEPENDENCE INCONSISTENTLY

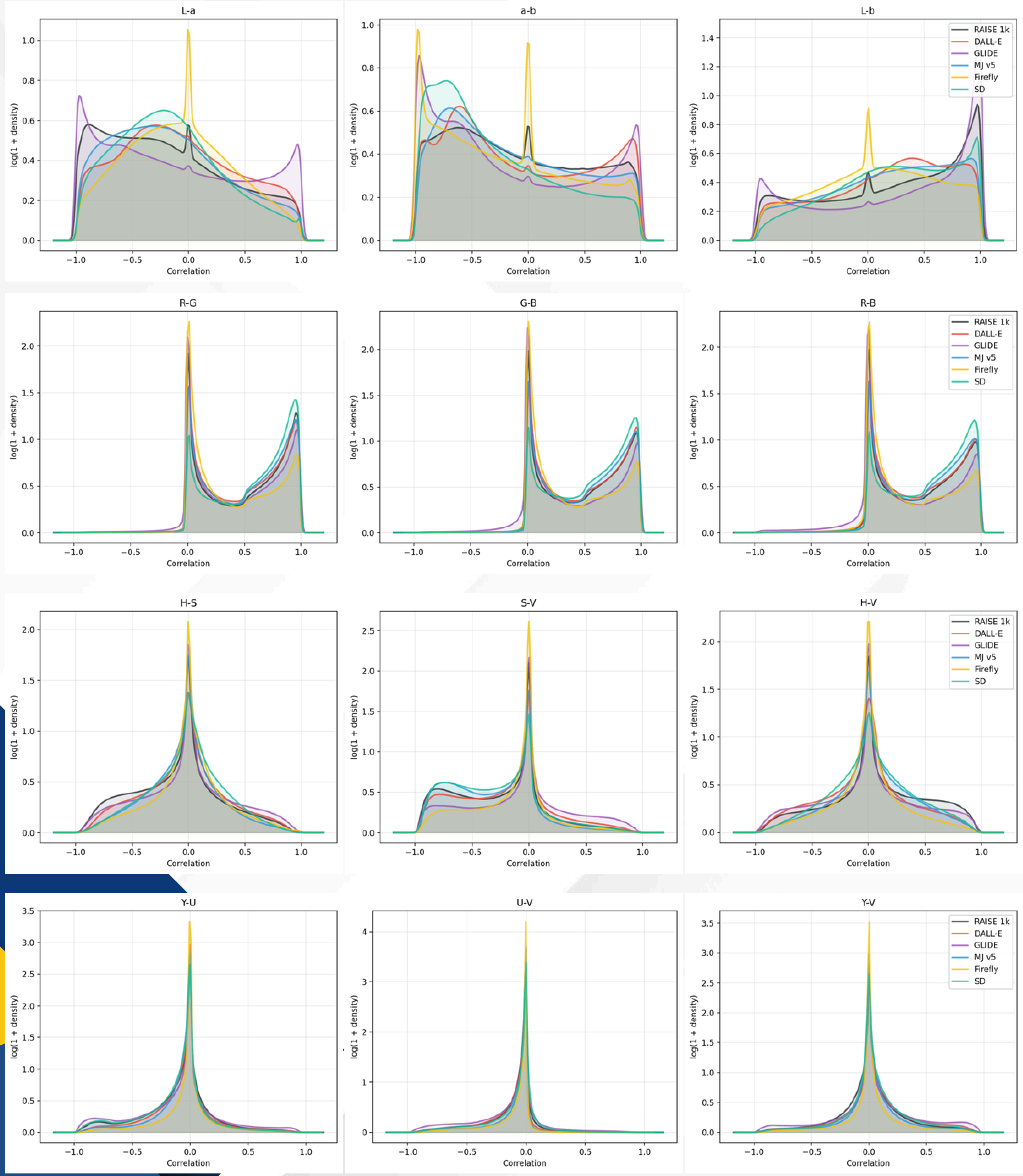
- YUV: large gap (Corr.>Uncorr.)
- RGB: large gap (Corr.>Uncorr.)
- Lab: gap nearly collapses
- HSV: large gap (UnCorr.>Corr.)

Equal RGB power → **very different LPIPS response**



INTER-CHANNEL DISTRIBUTIONS





Lab

RGB

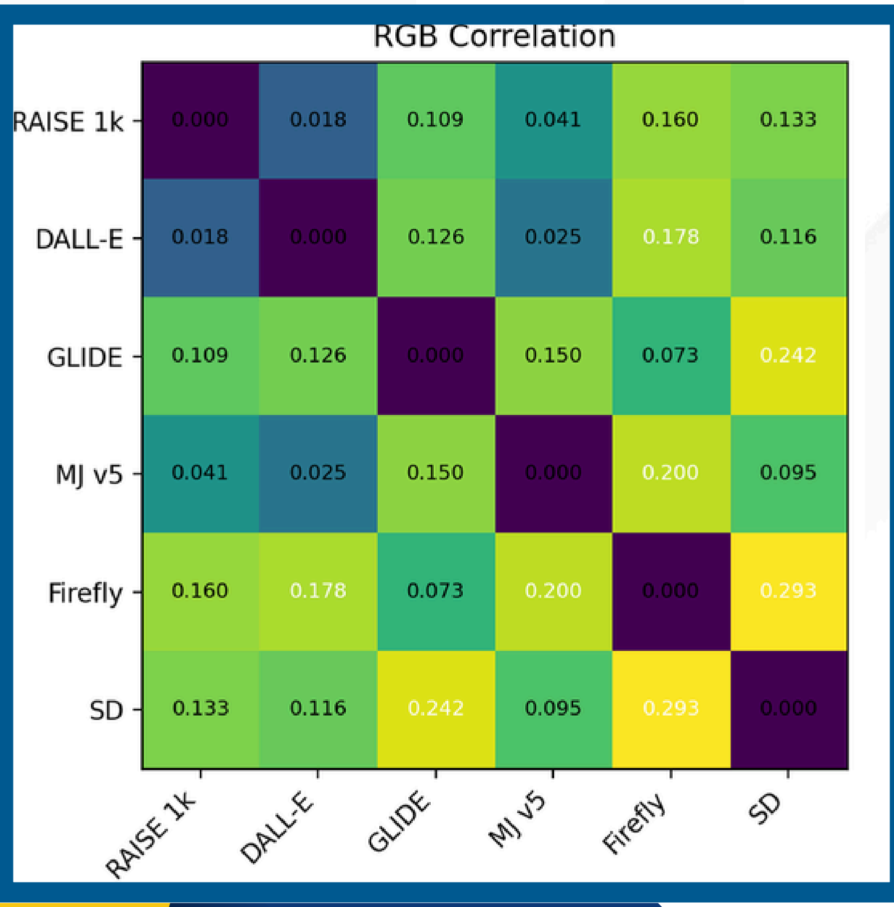
HSV

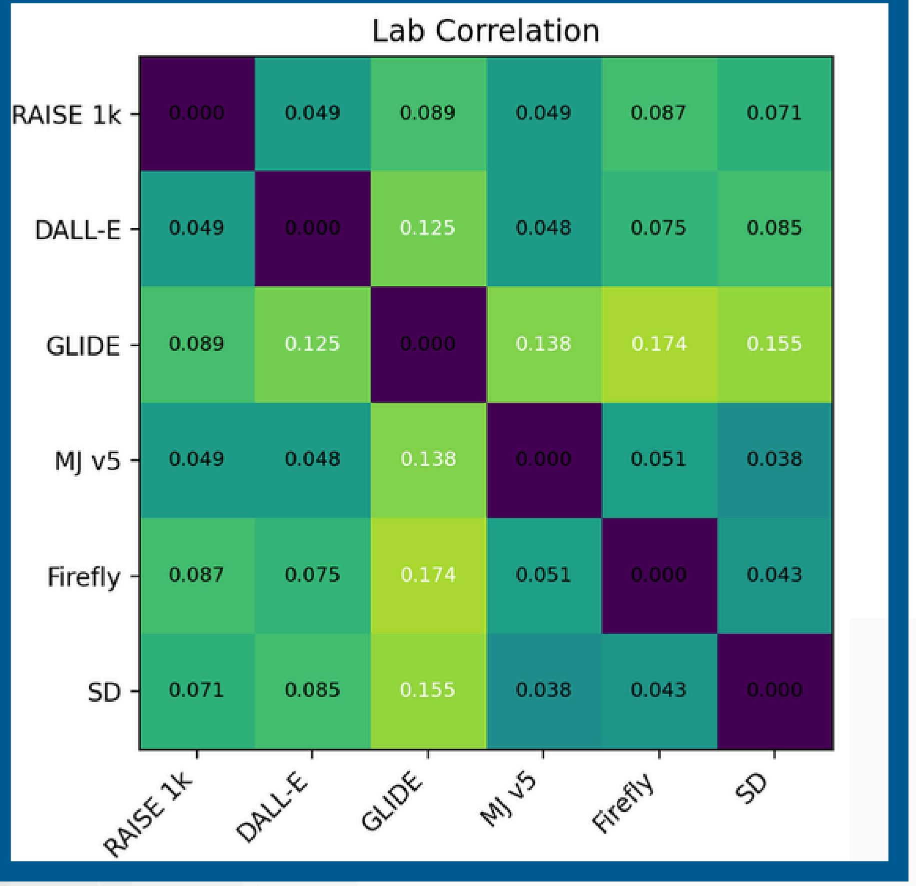
YUV

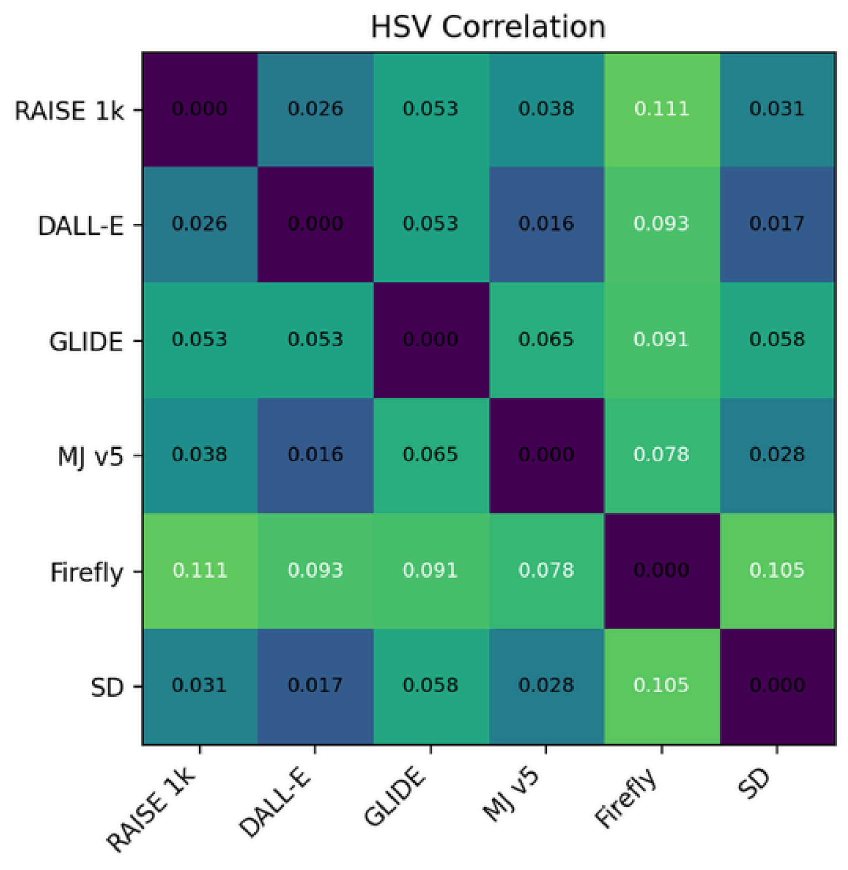
Differences in the distribution of real and generated images

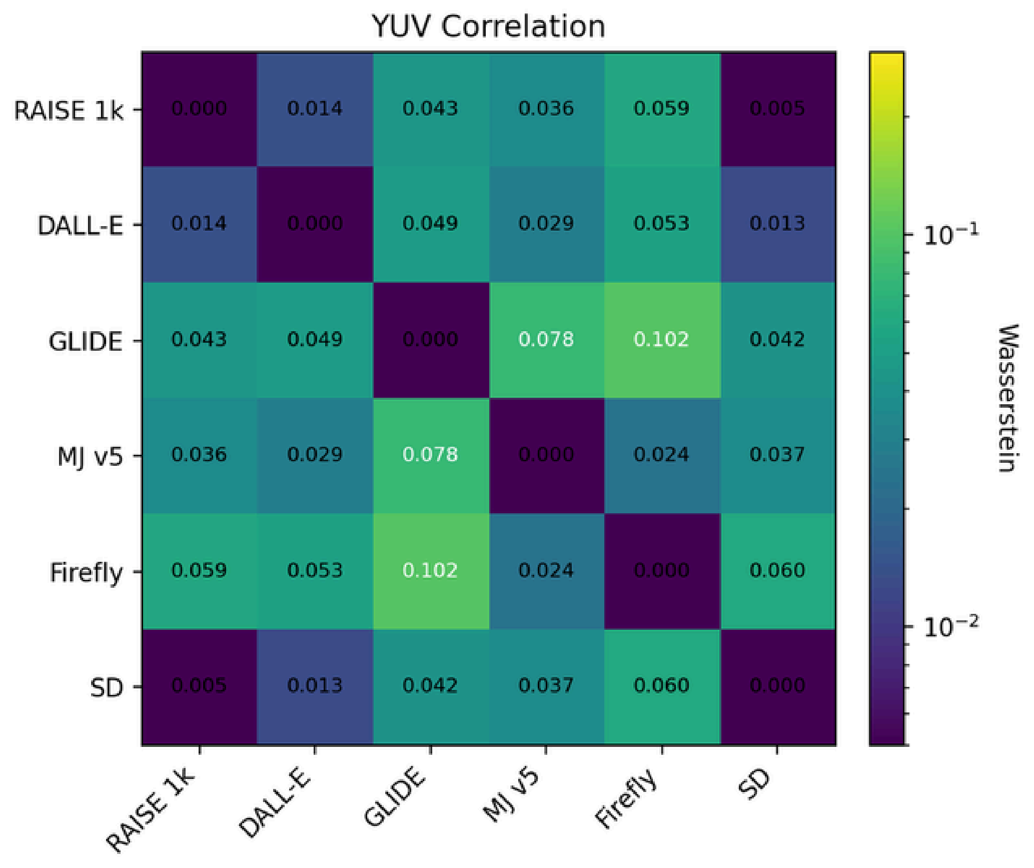
Differences in distribution among generators

WASSERSTEIN SEPARABILITY



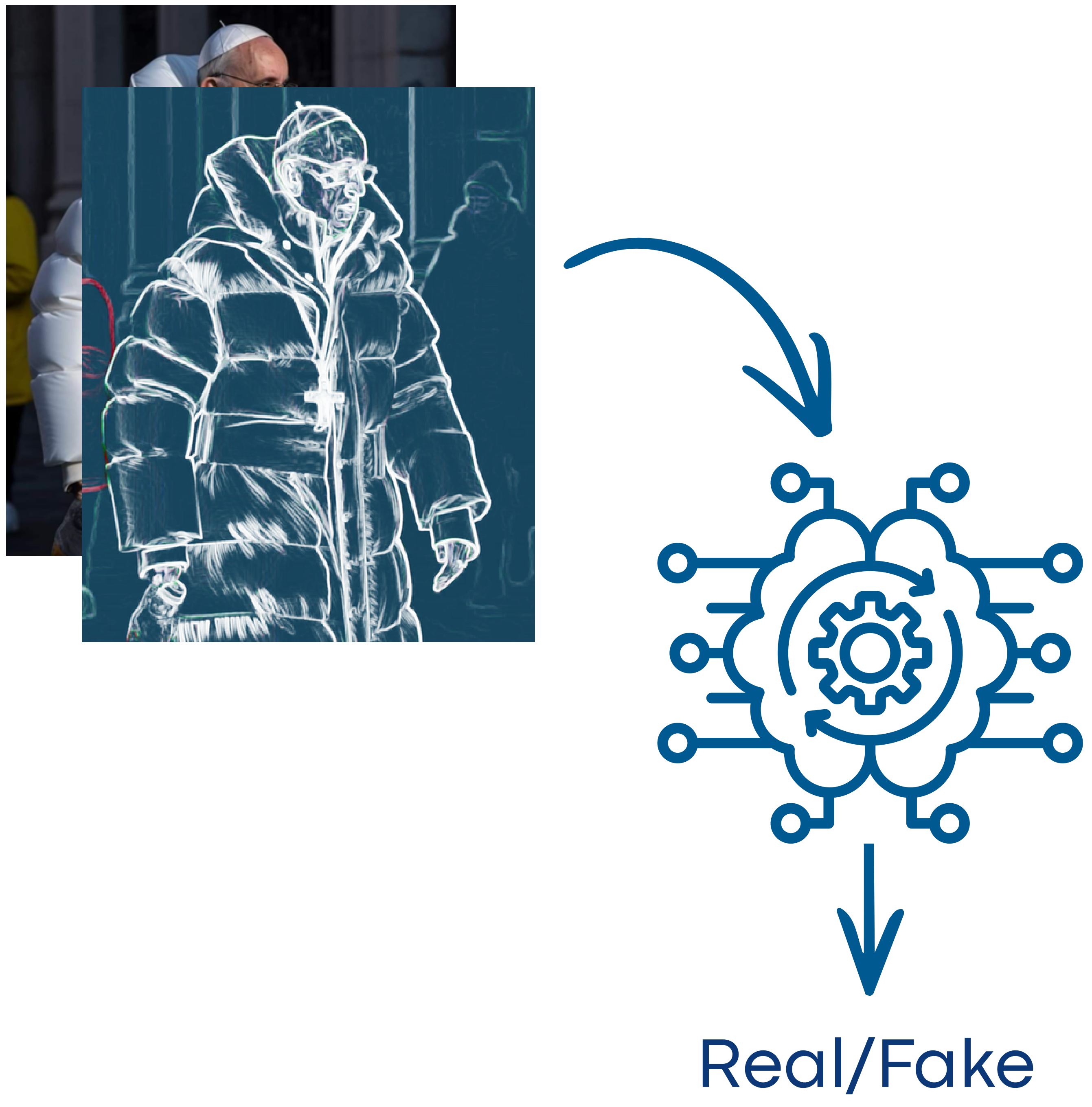






CHROMA: THE DETECTOR

- **Backbone**: ResNet-50, ImageNet-pretrained
- First conv layer adapted to the input channel count
- **Input variants** evaluated:
    - RGB (3 ch)
    - RGB + Corr(RGB) (6 ch)
    - RGB + Corr(Lab) (6 ch)
    - RGB + Corr(RGB) + Corr(Lab) (9 ch)
- **No new losses, no architectural tricks**, the contribution is the input representation



EVALUATION PROTOCOL

- **Benchmark of Cozzolino et al. (2024),** "Raising the Bar of AI-generated Image Detection with CLIP"
- **Test: 18 generators,** GAN family / diffusion family / commercial tools
- Two training regimes:
  - Single-generator: 10000 images from Latent Diffusion only
  - Multi-generator (†): 16088 images — +700 each from GLIDE and 5 Stable Diffusion variants, 588 StyleGAN2-FFHQ, 1 000 each DALL·E 2 / DALL·E 3
- Metric: AUC (%), cross-generator

Input variant	GAN family					Diffusion family									Commercial tools				AVG
	Pro GAN	Style GAN2	Style GAN3	Style GAN-T	Giga GAN	Score SDE	ADM	GLIDE	Latent Diff.	Stable Diff.	DeepFl. IF	Ediff-I	DiT	SDXL	DALL-E2	DALL-E3	Midj.	Adobe Firef.	
RGB	88.9	70.3	82.6	94.7	76.5	22.2	71.7	96.2	88.2	99.1	90.5	73.3	93.9	93.3	57.1	43.6	88.0	49.0	76.6
RGB + Corr(RGB)	65.1	47.6	51.6	10.5	54.1	16.9	58.3	66.2	62.8	87.5	79.8	59.3	70.5	69.9	42.3	**58.1**	49.9	**69.2**	56.6
RGB + Corr(Lab)	88.1	78.9	82.5	99.3	74.5	20.1	76.9	98.0	93.5	99.6	92.1	83.5	96.0	97.3	52.8	9.0	78.1	9.3	73.9
RGB + Corr(RGB) + Corr(Lab)	88.3	73.2	81.6	96.9	73.2	22.0	73.4	95.1	89.7	98.7	90.9	77.3	94.1	94.2	52.0	15.5	83.7	8.8	72.7
RGB†	91.0	68.5	58.0	99.9	77.5	96.8	81.0	100.0	94.4	100.0	97.6	59.3	**99.1**	100.0	100.0	37.1	**99.8**	36.9	83.2
RGB + Corr(RGB)†	89.4	71.8	59.2	99.5	71.7	92.1	81.4	99.8	93.8	100.0	97.6	65.8	94.7	100.0	100.0	44.5	98.6	30.0	82.8
RGB + Corr(Lab)†	96.1	89.4	86.5	**100.0**	82.7	**97.5**	**84.8**	99.6	97.4	99.9	**98.8**	**87.7**	98.3	100.0	100.0	20.2	97.7	3.8	**85.6**
RGB + Corr(RGB) + Corr(Lab)†	**97.4**	85.4	86.9	**100.0**	84.4	96.3	82.5	99.6	95.2	99.8	**98.8**	87.2	99.0	100.0	52.0	15.5	83.7	8.8	81.8
RGB + Corr(Lab Trainable)†	97.0	81.2	**87.4**	**100.0**	**85.0**	**97.5**	80.5	99.7	95.8	99.9	**98.9**	83.0	99.0	100.0	99.9	12.7	98.1	4.9	84.5
RGB + Corr (1x1 Trainable)†	89.8	74.8	64.5	99.1	77.6	90.8	82.5	99.9	95.0	100.0	97.3	64.5	96.8	100.0	99.9	25.7	96.6	21.5	82.0

Table 4.2: Ablation over input representations (AUC, %) under the benchmark protocol of Cozzolino et al. [16]. † indicates training with additional generator data; gray marks in-domain generators. **Bold** and underline denote the best and second-best results per column.

LAB CORRELATIONS HELP

- With enough generator diversity, Lab correlations add **+2.4 pts over RGB**
- **RGB correlations hurt**, they're largely redundant with the RGB input itself
- Under **single-generator training**, correlation features **overfit the source generator**

Input variant	**Single-generator**	**Multi-generator †**
RGB	76,6	**83,2**
RGB + Corr(RGB)	56,6	82,8
RGB + Corr(Lab)	73,9	**85,6**
RGB + Corr(RGB) + Corr(Lab)	72,7	81,8

avg AUC (%) over 18 generators

Method	GAN family					Diffusion family									Commercial tools				AVG
	Pro GAN	Style GAN2	Style GAN3	Style GAN-T	Giga GAN	Score SDE	ADM	GLIDE	Latent Diff.	Stable Diff.	DeepFl. IF	Ediff-I	DiT	SDXL	DALL-E2	DALL-E3	Midj.	Adobe Firef.	
Wang et al.	100	96.5	98.5	98.9	66.6	32.9	64.3	48.5	59.2	41.5	78.0	64.9	58.6	54.3	64.8	10.9	40.2	84.8	64.6
PatchFor.	92.3	84.5	91.8	91.2	64.7	83.3	74.8	96.2	78.1	62.4	62.7	78.7	83.1	68.4	41.9	52.7	57.8	49.4	73.0
Grag. et al.	100	**99.8**	97.5	98.8	82.8	92.1	74.7	62.8	91.9	52.5	69.9	69.6	65.3	58.0	58.3	2.4	43.1	63.5	71.3
Mand. et al.	96.2	93.8	**100.0**	92.6	61.8	99.8	56.5	40.5	70.0	36.8	47.2	65.0	59.1	27.0	14.5	14.7	24.3	36.7	57.6
Liu et al.	100	**99.8**	98.4	98.5	**98.2**	95.4	82.5	76.5	97.6	77.4	72.2	98.7	88.0	31.1	70.4	0.2	40.7	11.8	74.3
Corvi et al.	79.4	73.7	50.0	97.1	63.4	65.0	80.7	91.9	100.0	**100.0**	**99.9**	85.7	**100.0**	**100.0**	69.4	60.8	**100.0**	**98.0**	84.2
LGrad	100	91.2	83.8	81.8	82.2	80.6	76.9	66.1	81.1	61.5	68.8	74.1	56.2	57.2	58.6	37.9	56.3	40.6	69.7
Ojha et al.	100	93.9	92.3	98.2	96.0	58.4	**86.7**	80.8	85.7	89.5	92.9	80.6	77.8	85.1	95.2	36.4	66.2	97.5	84.1
DIRE-1	50.6	56.9	47.8	99.9	74.1	44.3	75.7	71.4	68.7	39.4	98.9	99.1	99.6	47.1	44.7	47.6	51.0	57.4	65.2
DIRE-2	54.2	52.5	43.0	99.6	76.0	41.0	70.1	70.1	69.3	46.9	97.0	98.2	98.3	42.8	41.0	49.6	47.8	43.0	63.3
NPR	100	85.6	77.0	96.4	88.7	91.1	86.3	79.3	90.2	64.5	91.6	80.1	78.4	76.7	39.5	48.7	77.0	32.1	76.8
Cozzolino et al. 10k+	93.4	87.1	87.6	99.9	78.5	89.2	79.9	**99.7**	84.7	91.3	97.9	**99.4**	94.0	90.1	86.3	**92.9**	81.7	87.2	**90.0**
RGB + Corr(Lab)†	96.1	89.4	86.5	**100.0**	82.7	97.5	84.8	99.6	97.4	99.9	98.8	87.7	98.3	100.0	100.0	20.2	97.7	3.8	85.6
RGB + Corr(RGB) + Corr(Lab)†	97.4	85.4	86.9	**100.0**	84.4	96.3	82.5	99.6	95.2	99.8	98.8	87.2	99.0	100.0	52.0	15.5	83.7	8.8	81.8
RGB + Corr(Lab Trainable)†	97.0	81.2	87.4	**100.0**	85.0	97.5	80.5	99.7	95.8	99.9	98.9	83.0	99.0	100.0	99.9	12.7	98.1	4.9	84.5

Table 4.3: Comparison with prior methods in terms of AUC (%) under the benchmark protocol of Cozzolino et al. [16]. Results are reported across GAN-based generators, diffusion-based generators, and commercial tools. Gray entries indicate the in-domain generator(s) associated with each method's training setup as reported by the protocol, whereas the remaining columns measure cross-source generalization. **Bold** marks the best overall average and underline marks the second-best.

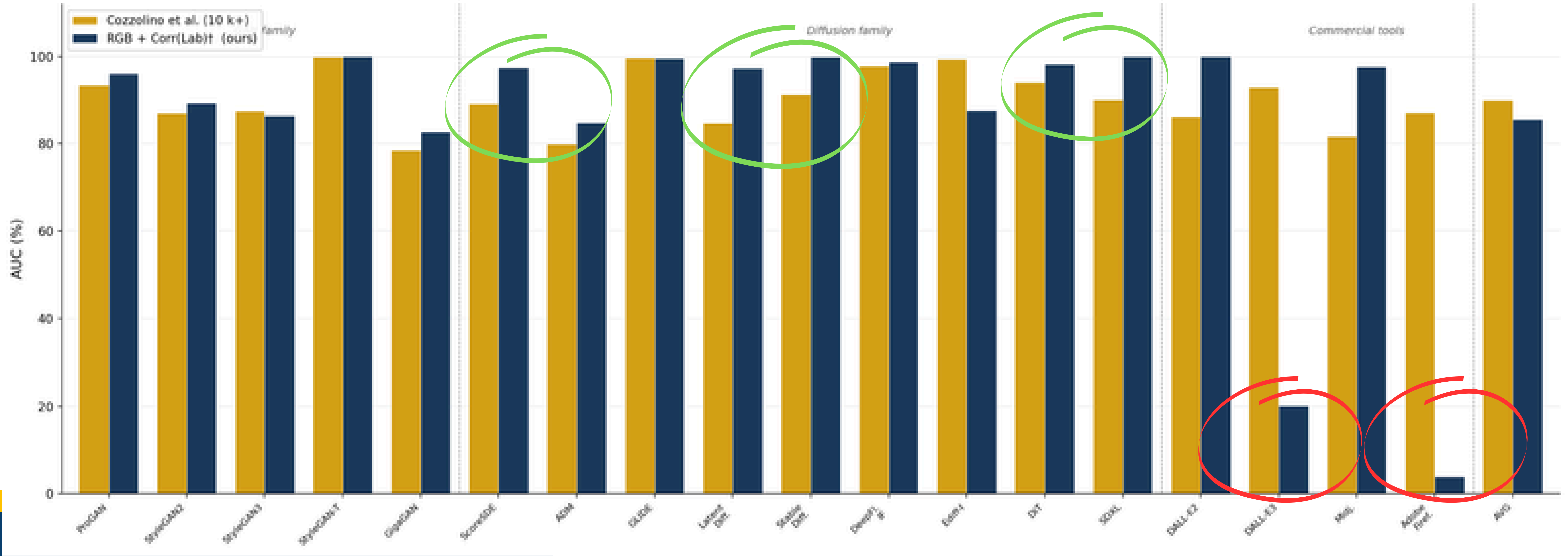
COMPETITIVE WITH FAR HEAVIER DETECTORS

Cozzolino

90.0%

Ours

85.6%



CONCLUSIONS

- **Takeaways**
  - LPIPS **does not enforce channel dependence** uniformly across color parameterizations
  - Generated images **alter inter-channel relationships**, and the effect is representation-dependent
  - **Correlation descriptors are complementary to RGB**, not a replacement: **+2.4 AUC** on top of a stock ResNet-50
- **Limitations**
  - **Heavy, unknown post-processing destroys the correlation structure**
  - Gains require **generator diversity at training time**
- **Future Work**
  - Learn a non-linear forensic color space · fuse with spectral cues · extend to localization

THANK YOU!

Scan Me

Scan Me

Scan Me

UNIVERSIDAD DE LA REPUBLICA URUGUAY

FACULTAD DE INGENIERÍA

école normale supérieure paris—saclay

CENTRE BORELLI