

Semantic Color Naturalness Breaker: Preventing Illegitimate Colorization via Content-Aware Color Priors

Yuki Nii¹, Futa Waseda^{1,2}, Ching-Chun Chang¹, Isao Echizen^{1,2}

National Institute of Informatics, The University of Tokyo



Motivation: Recent Threats

Advancements of colorization models can pose **copyright infringements**.

Unauthorized AI colorization of manga and illustrations are sold online.
We need to protecting the copyrighted contents!

I will color and repaint manga and comic pages using ai

Max Y Level 2 ♦♦♦ 1 order in queue
★★★★★ 5.0 (61 reviews)

Basic	Standard	Premium
	Colors and shading (2 PAGES)	US\$15
i will color , shade , texture your comic , manga page using a costume AI model i trained		
🕒 3-day delivery 🔄 3 Revisions		
✓ 1 prompt ✓ Prompt creation ✓ Prompt delivery ✓ Sample images ✓ Instructions ✓ Artwork delivery		
Continue		
Contact me		

I will color manga, comic, webtoon and illustration

Giuseppefb Level 2 ♦♦♦ 1 order in queue
★★★★★ 4.9 (284 reviews)

I will color your manga, comic or illustration

Vira Level 2 ♦♦♦ 1 order in queue
★★★★★ 5.0 (128 reviews)

[3] <https://www.fiverr.com/striderxyz/color-and-repaint-manga-and-comic-pages-using-ai>

Motivation: Recent Threats

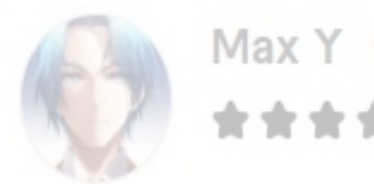
Advancements of colorization models can pose copyright infringements.

Unauthorized AI colorization of manga and illustrations are sold online.
We need

**Can we protect grayscale images
before they are shared?**

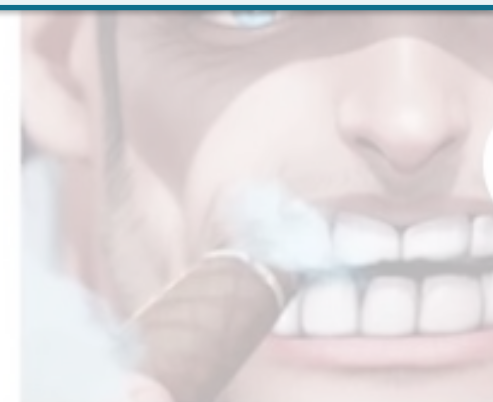
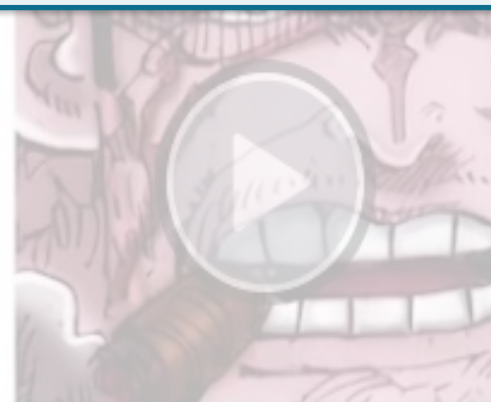
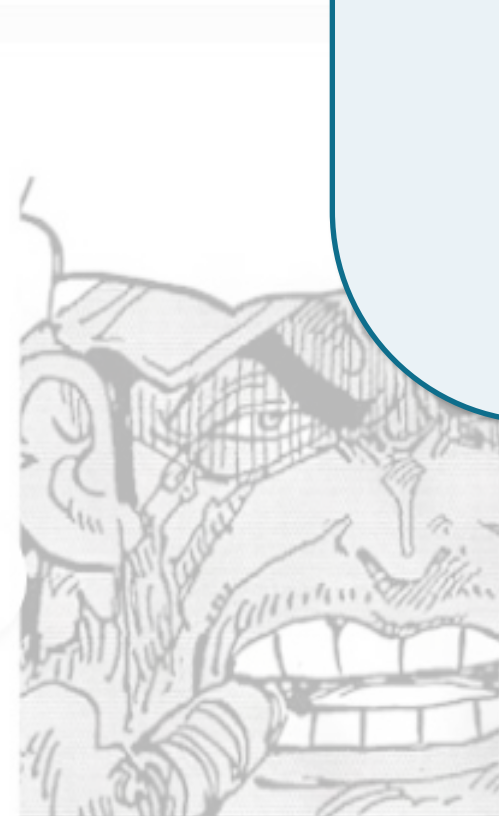
**=> We can invalidate later colorization by
leveraging adversarial noise!**

I will color
using ai



Max Y

★★★★☆



✓ Artwork delivery

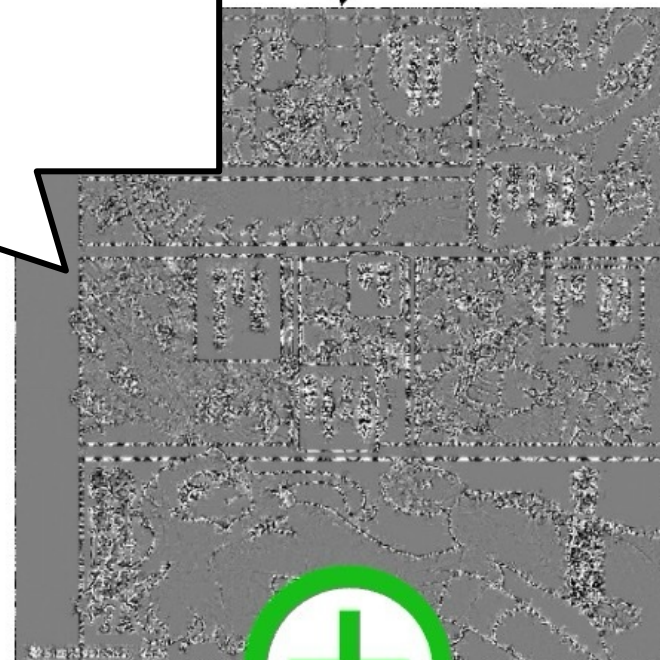
Continue →

Contact me



Prior Work : PACHroma [Nii et al. 2025]

**Adversarial
noise** to disrupt
colorization
model



Unprotected
Manga



Imperceptible
Perturbation

Uncolorable
Example

Illegitimate AI
colorization

Illegitimate AI
colorization

×
Copyright Infringed



✓
Copyright Protected

PACHroma aims
to **suppress color**
in the output.

Proactive defense!

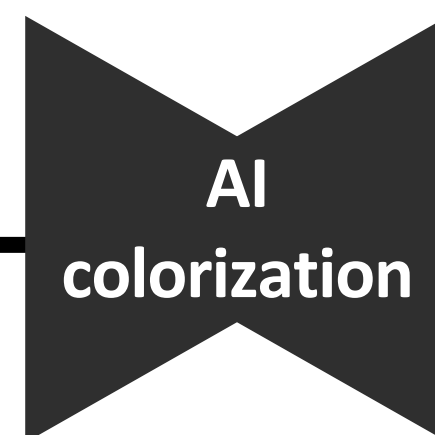


But is **Desaturation** enough?

Not really. Grayish can still look natural.



UE PACHroma



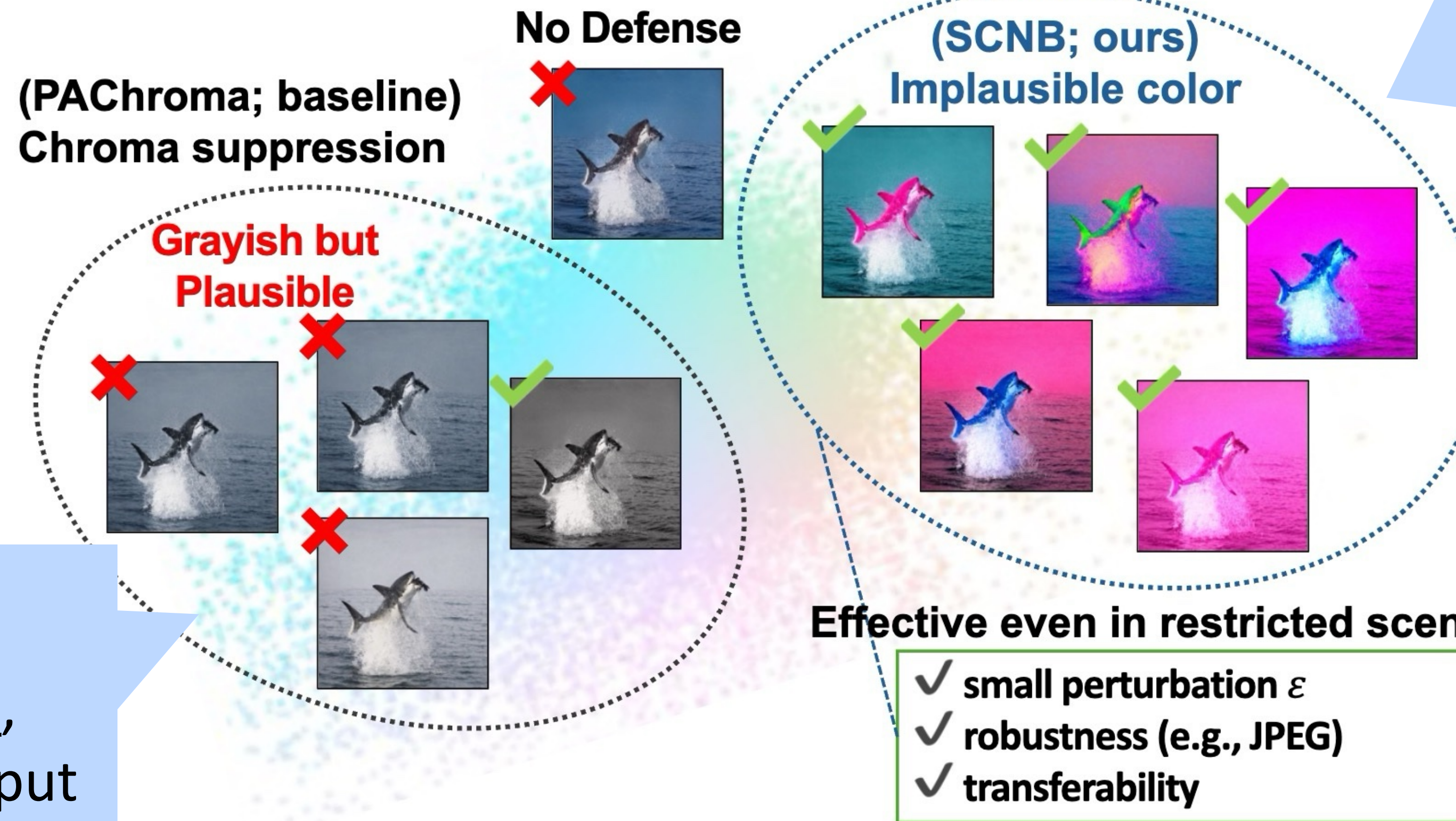
Colored Output

Cloudy day?
Rock with algae?
(**Plausible**)

Our Idea: Push Toward Unnatural Colors

We aim to push the colored output towards unnatural colors!

Even when complete desaturation fails, the image will look implausible.

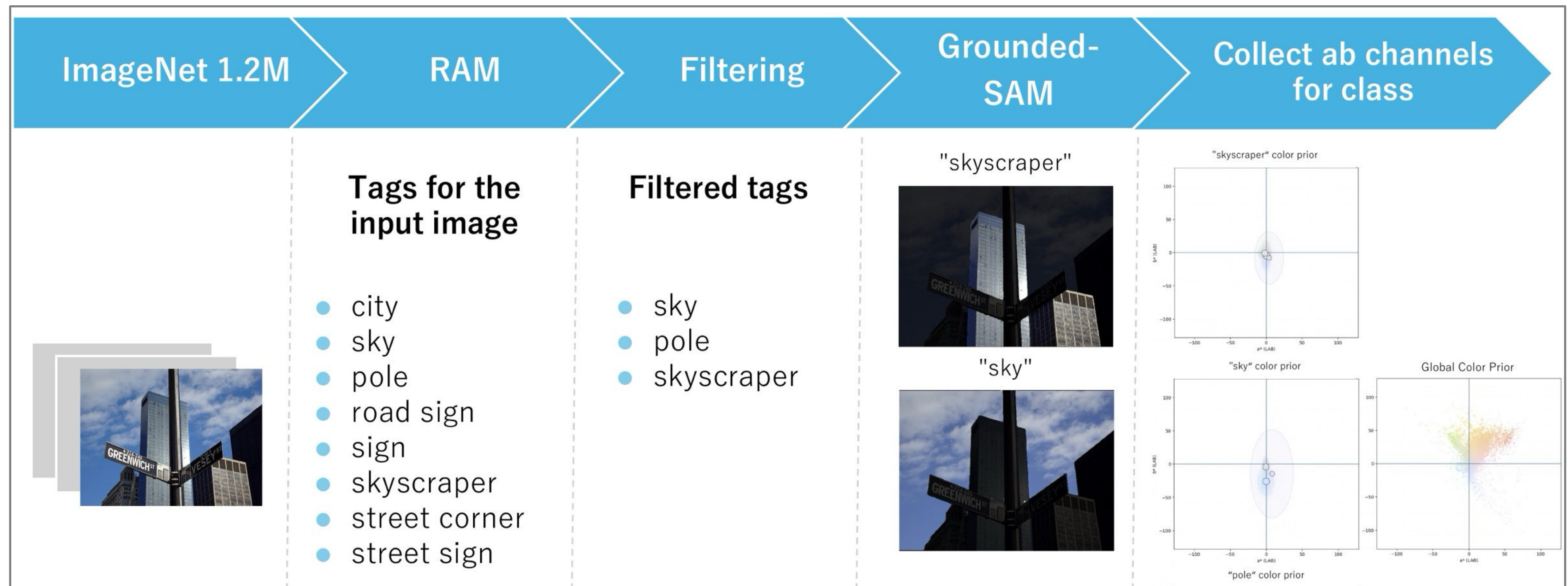


When complete desaturation fails, the colored output can look plausible.

How to Measure “Naturalness” of the Colors?

→ CaCDD: Content-Aware Color Distributional Distance

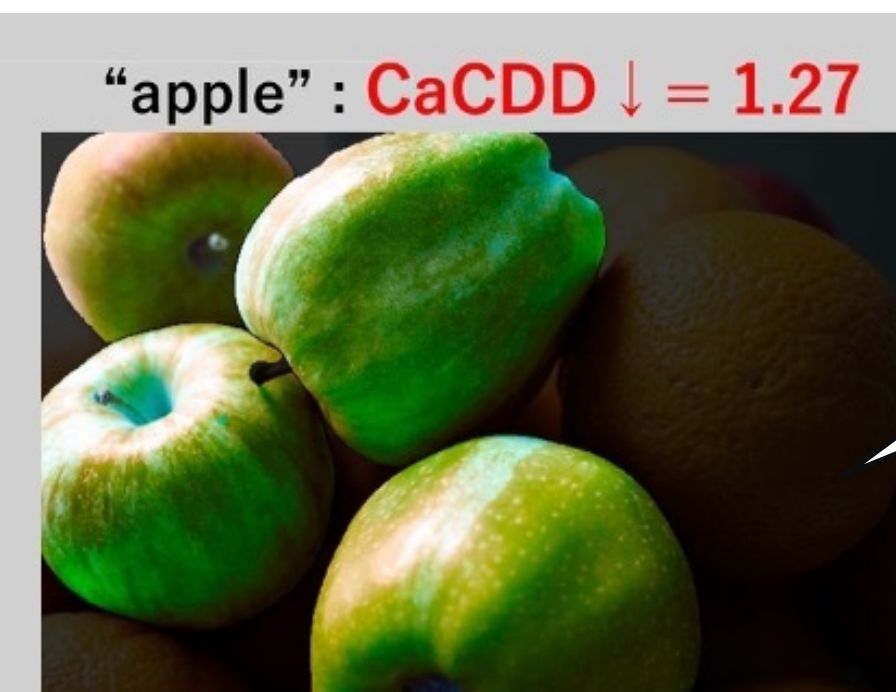
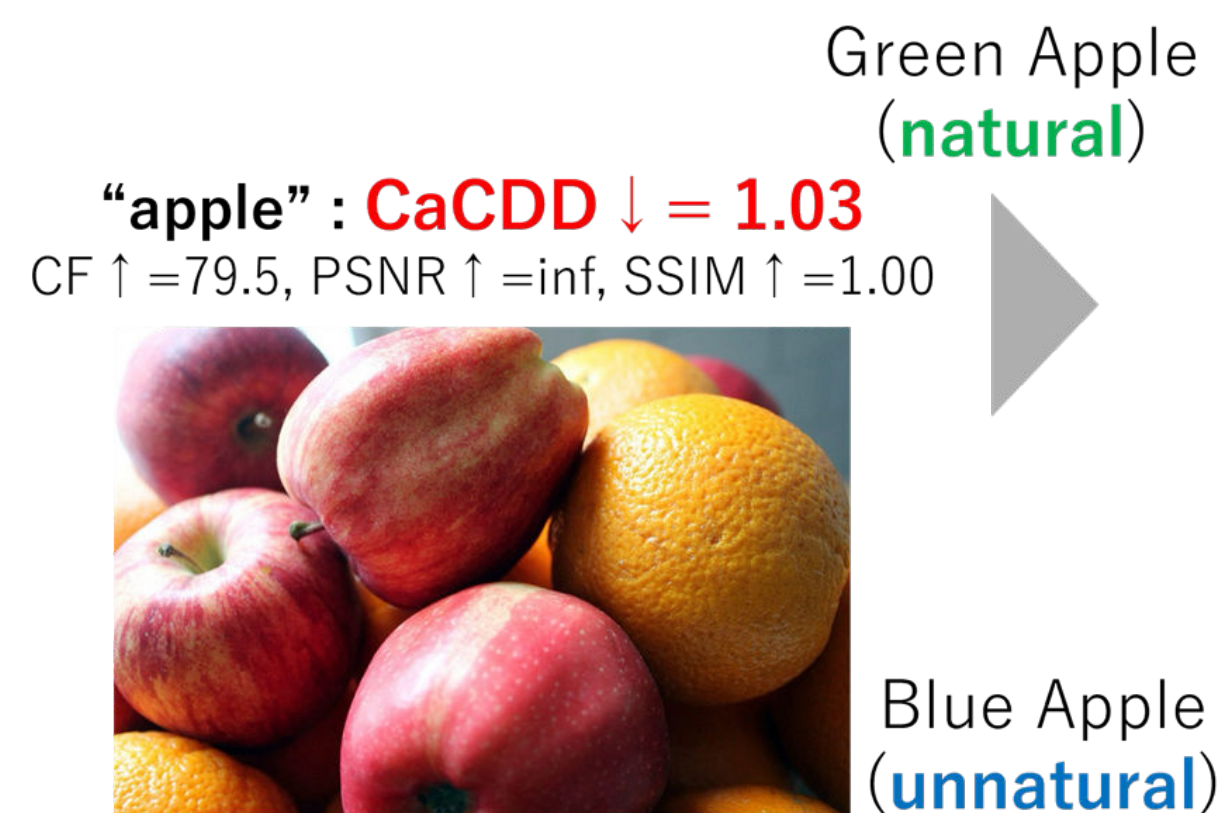
(1) We first collect chroma priors (GMM) in Lab ab-plane from large data for each class



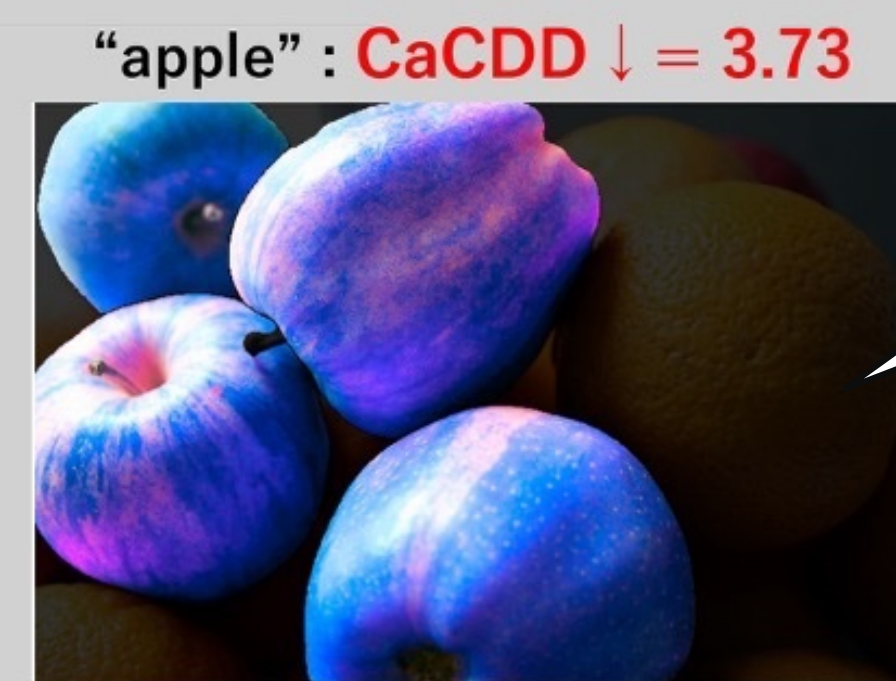
How to Measure “Naturalness” of the Colors?

→ CaCDD: Content-Aware Color Distributional Distance

(2) Then, we can measure the “unnaturalness” based on the distance from the obtained distribution



Different color,
but **natural**



Different color,
and **unnatural**

CaCDD: Content-Aware Color Distributional Distance

CaCDD quantifies how *semantically plausible* the color is **without needing GT**.

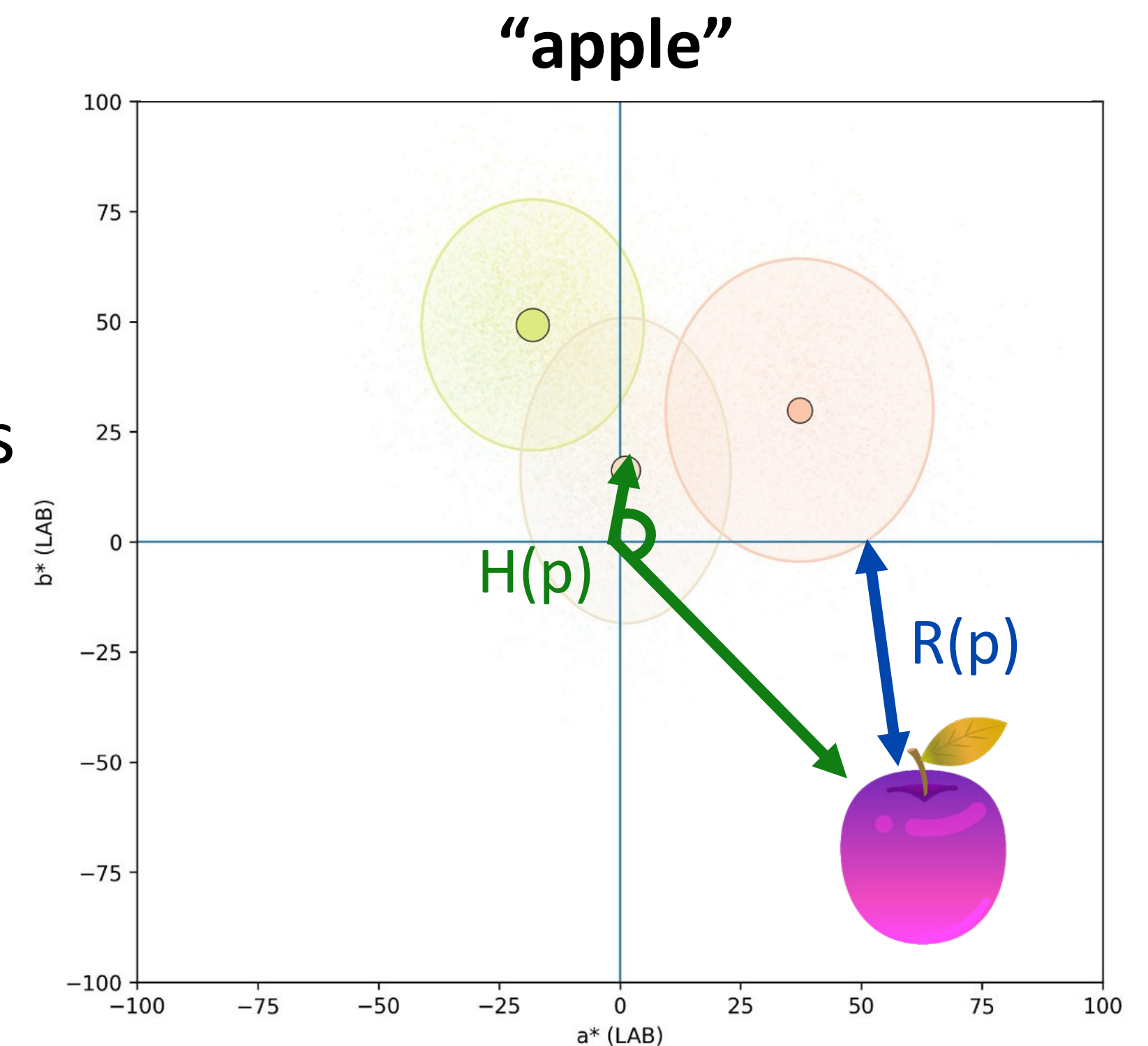
$$\ell_{\text{CaCDD}}(p) = R(p) + \lambda_{\text{hue}} w_p H(p)$$

R(p) : Responsibility-weighted Repulsion

how far the pixel's chroma is from the closest plausible colors using a **Mahalanobis distance** to GMM components

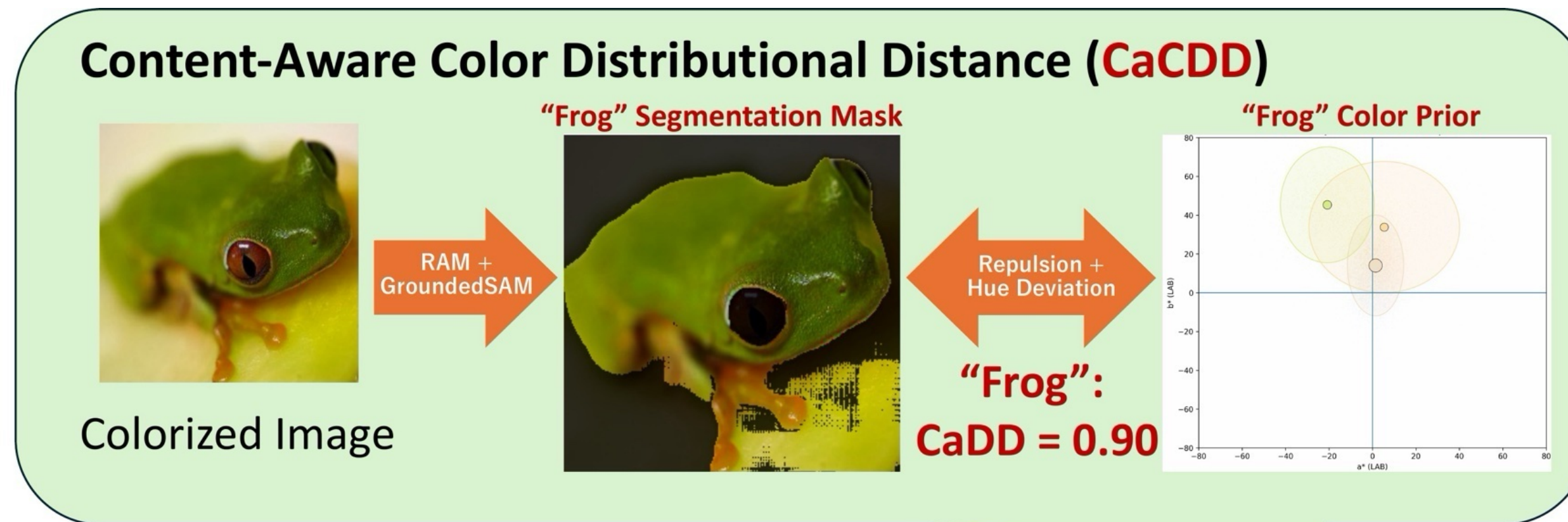
H(p) : Hue-direction deviation

measures whether the chroma points is in an **implausible hue direction (angle)**

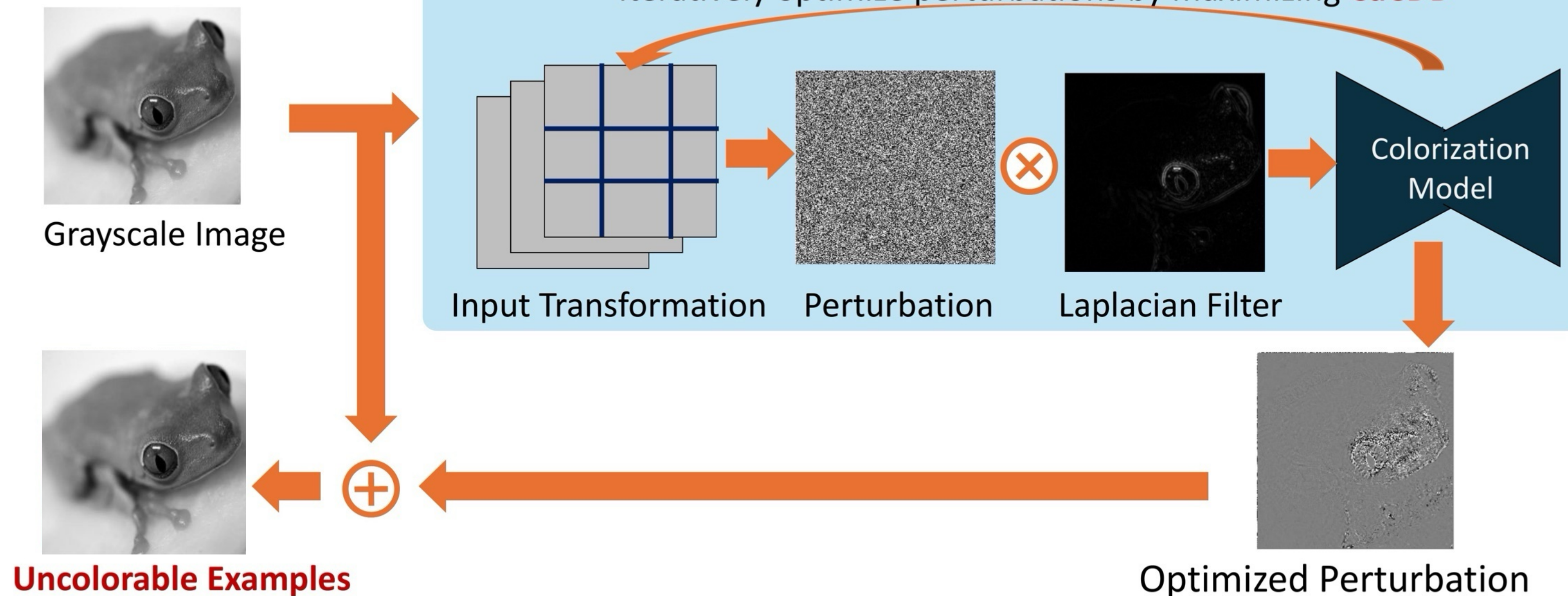


Semantic Color Naturalness Breaker (SCNB)

Maximize CaCDD while
keeping the published
input visually similar

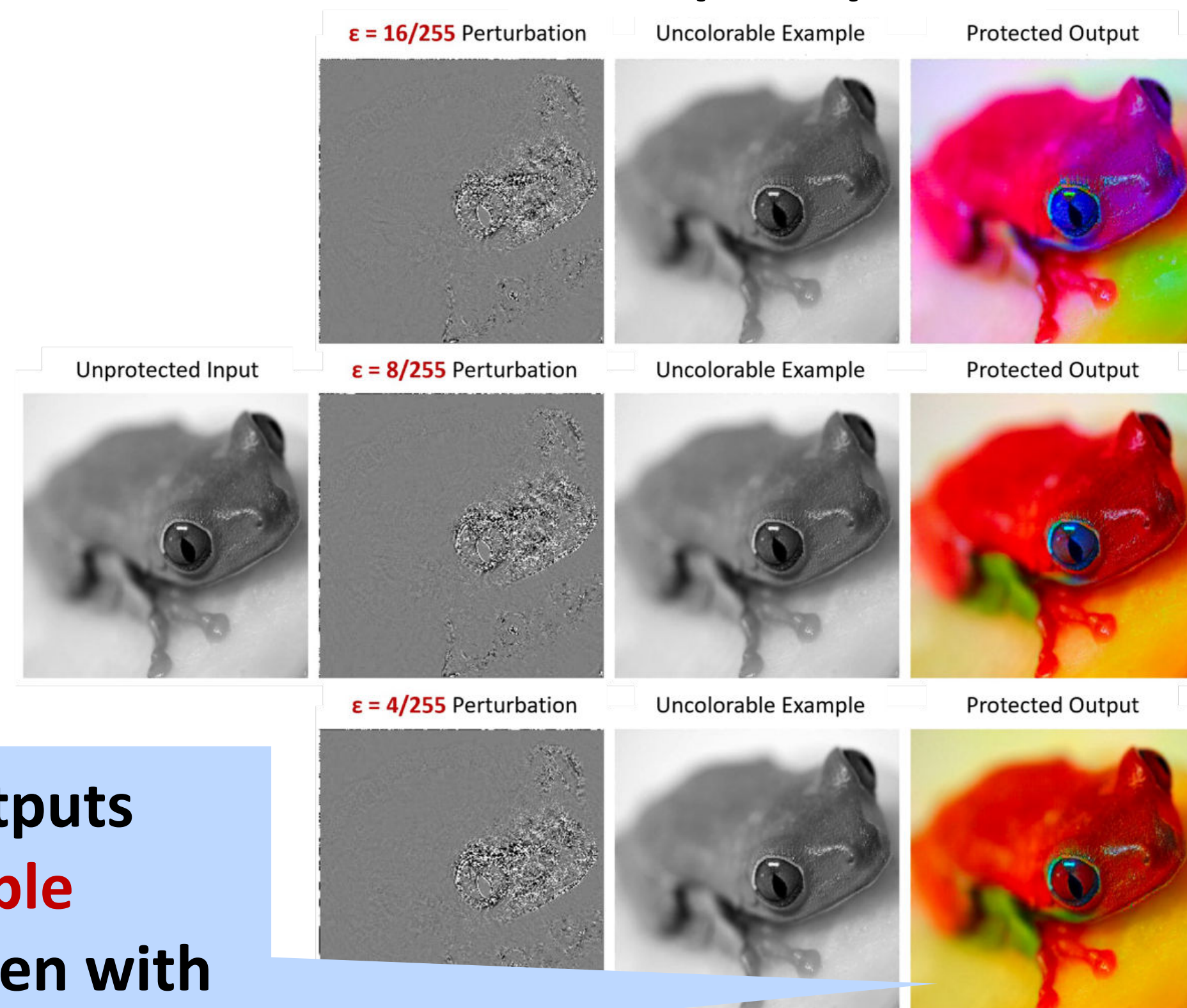


Proactive Defense



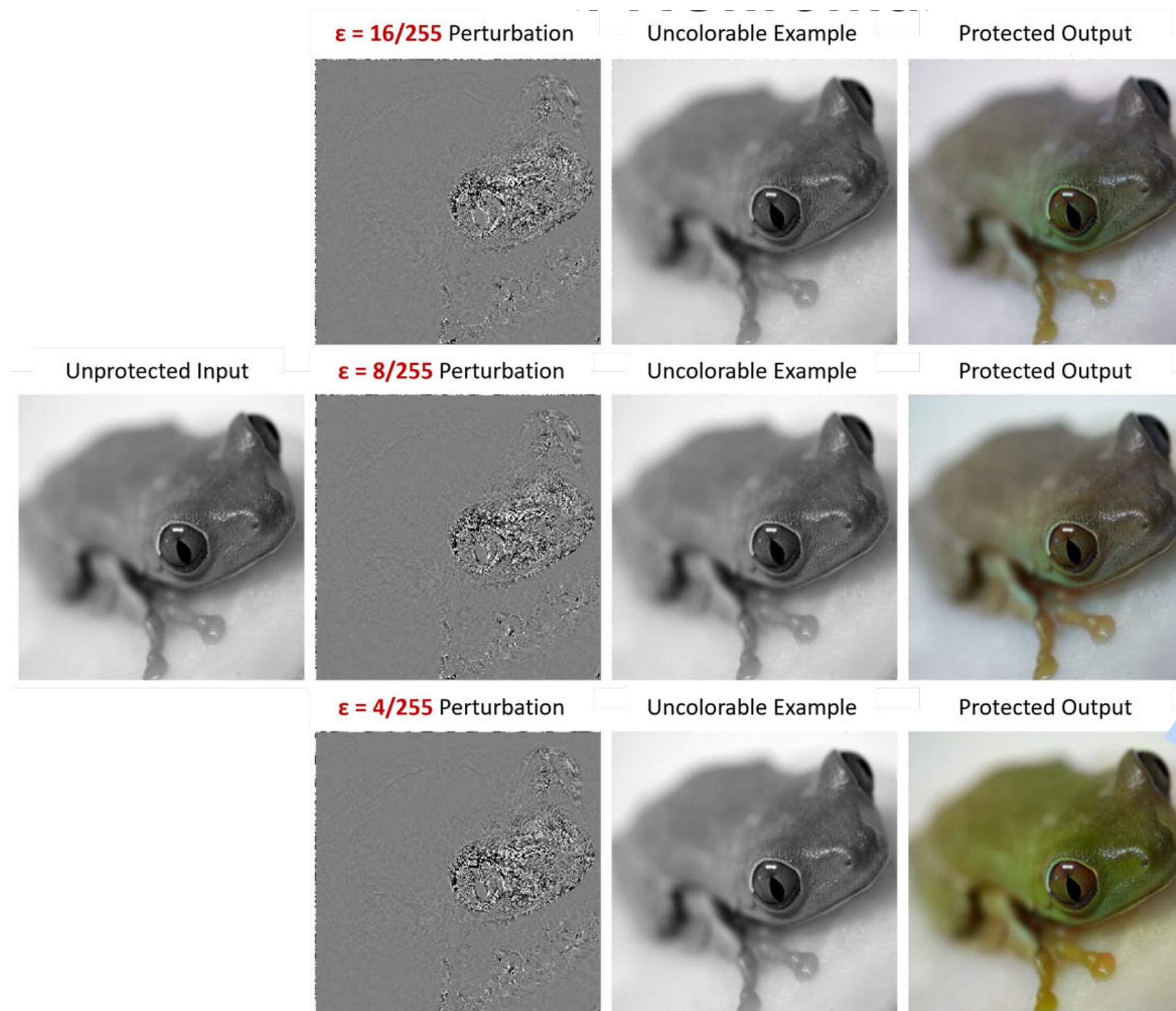
Effectiveness under smaller perturbation budget

SCNB (ours)



SCNB outputs
implausible
colors even with
 weaker noise

PAChroma



PAChroma
 outputs
desaturated but
plausible colors
 with weaker
 noise

Robustness to post-processing

SCNB (ours)

PAChroma

Uncolorable Example

Protected Output

Clean Output



Uncolorable Example

Protected Output

Clean Output



SCNB
(ours)

Jpeg75 Input

Jpeg75 Output

Jpeg50 Input

Jpeg50 Output

RRC Input

RRC output



PAChroma
(ours)

Jpeg75 Input

Jpeg75 Output

Jpeg50 Input

Jpeg50 Output

RRC Input

RRC output



SCNB outputs **implausible colors** even after common post-processing

PAChroma outputs **desaturated but plausible colors** after common post-processing

Transferability Across Unseen Colorizers

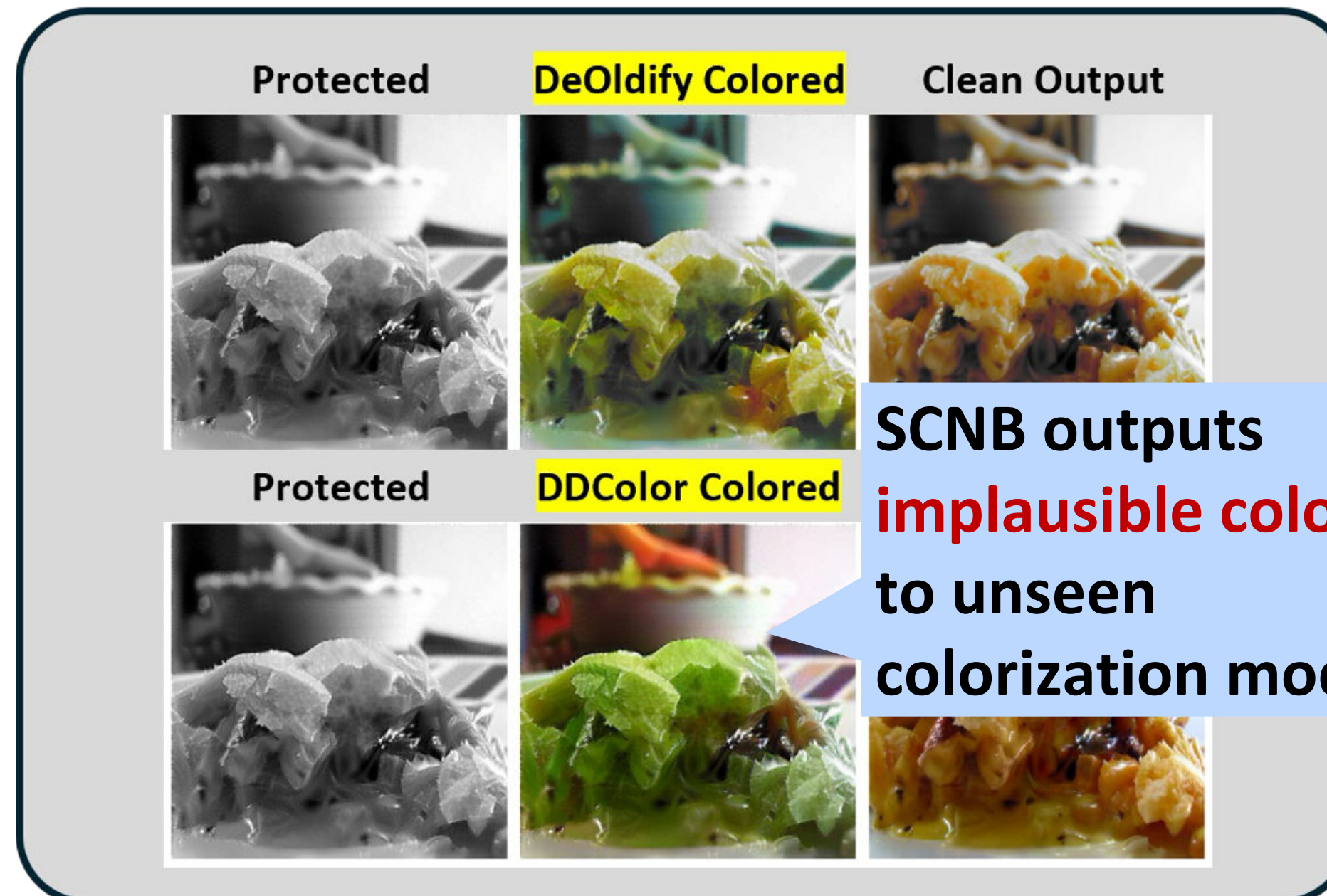
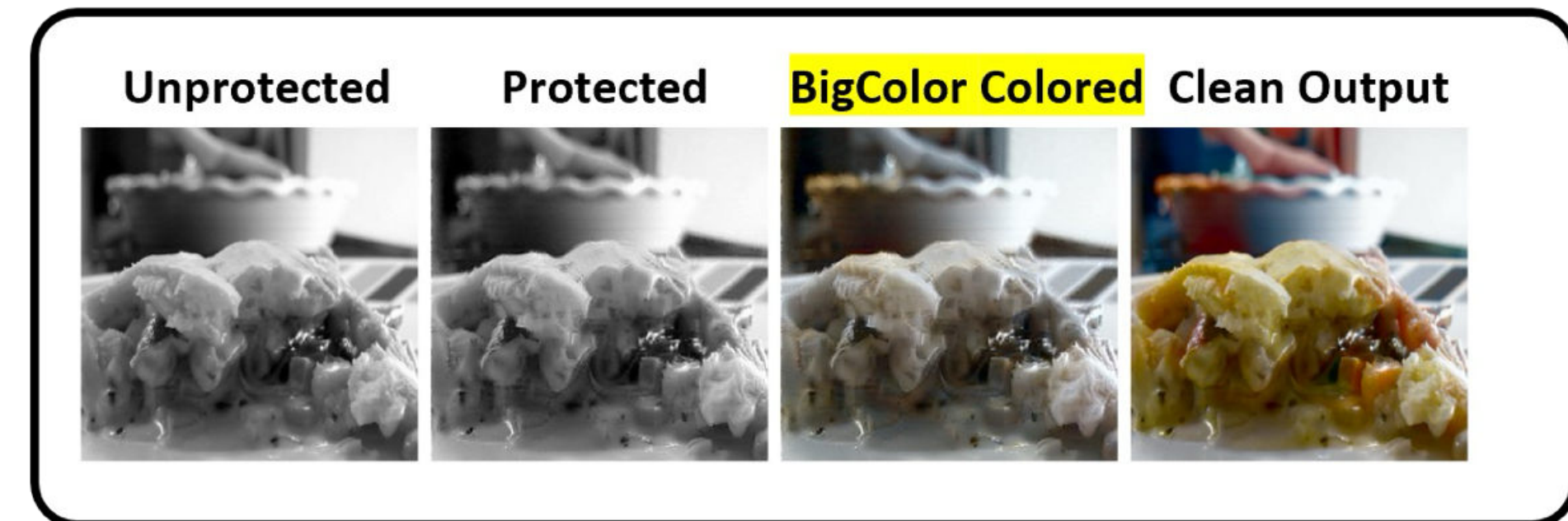
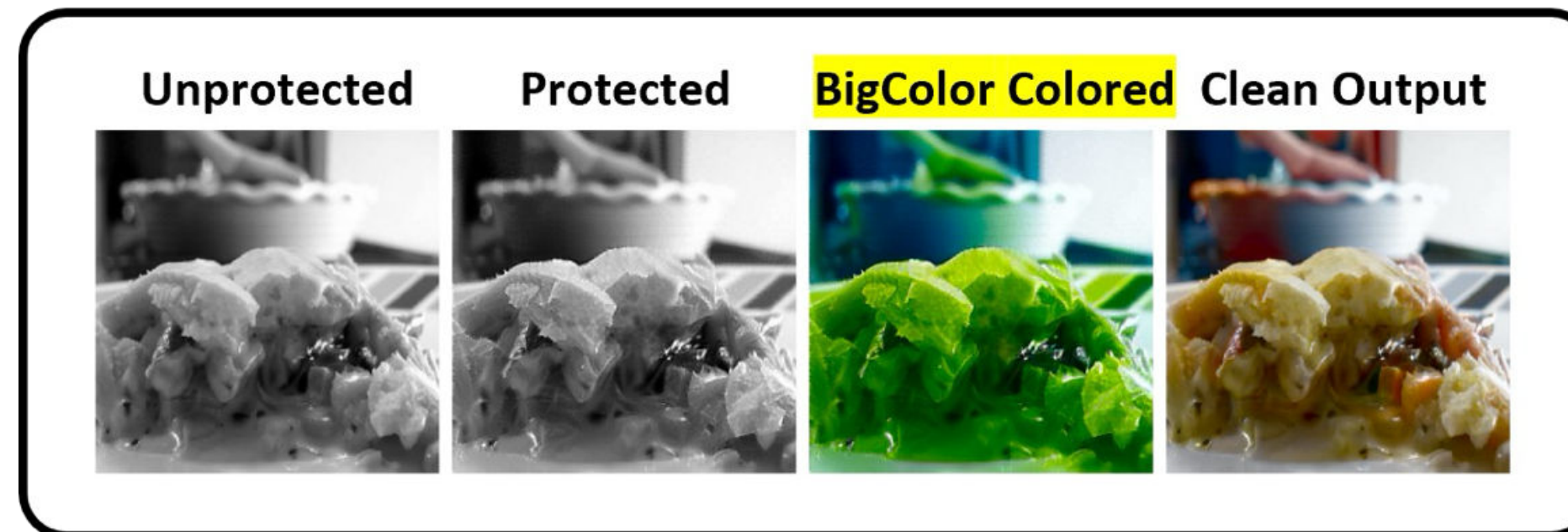
Source Model : BigColor

White box

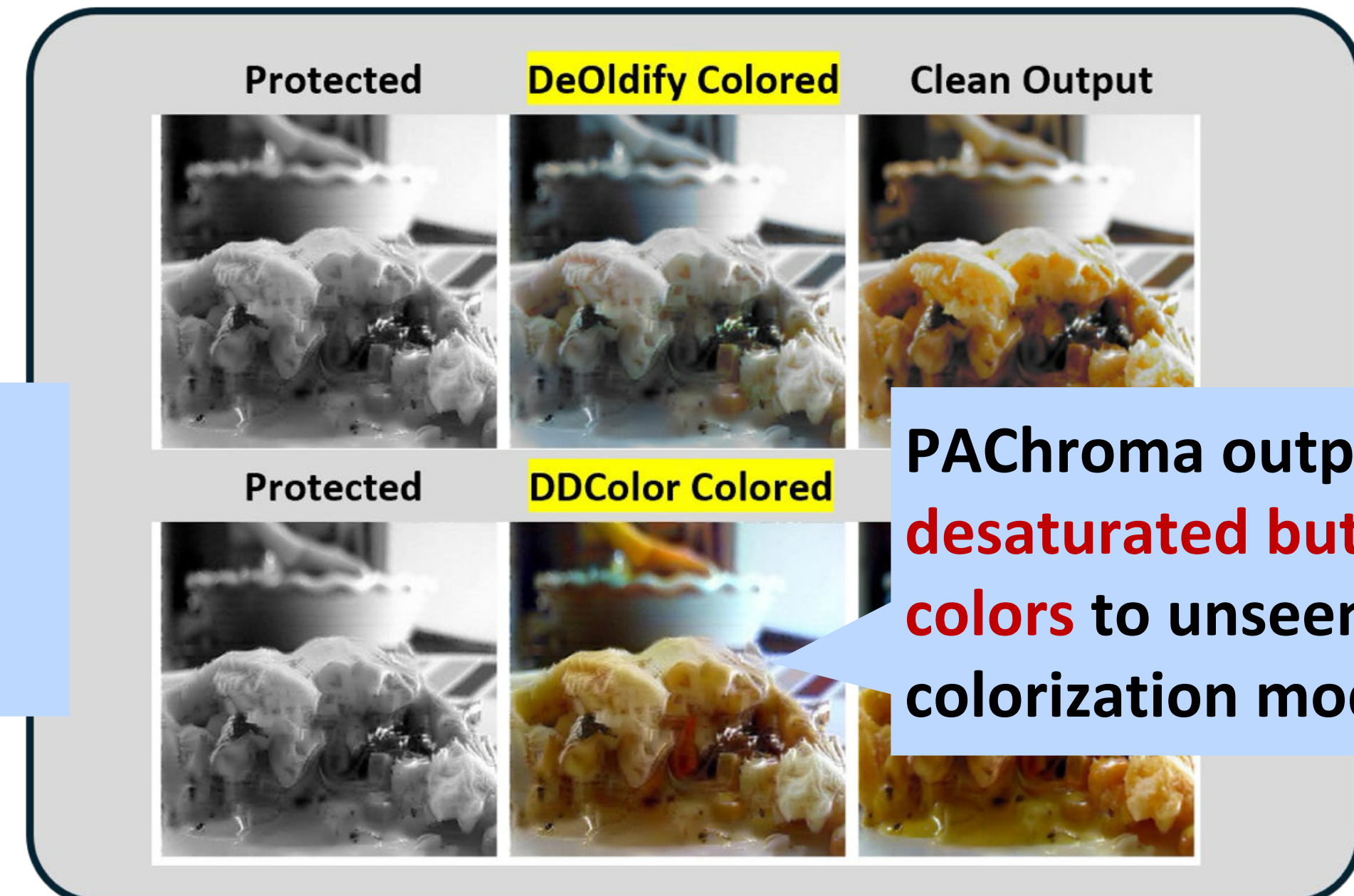
Black box

SCNB (ours)

PAChroma



SCNB outputs **implausible colors** to unseen colorization model



PAChroma outputs **desaturated but plausible colors** to unseen colorization model

Conclusion

- **SCNB makes colorized outputs semantically implausible.**
- **CaCDD provides the objective for generating protected images.**
- **The effect persists with smaller perturbations, common post-processing, and black-box transferability experiments.**