Semantic Object Priors for Robust Color Constancy in Indoor Scenes

By
Gautham Ramkumar
Sai Vamsi Rithvik Allanka
Yoga Srinivas Reddy Kasireddy

PROBLEM STATEMENT

Human vision naturally adjusts to different lighting conditions, perceiving a white refrigerator as white whether under warm tungsten bulbs or cool daylight. However, cameras capture the actual light reflected from surfaces, resulting in unwanted color casts.

The Core Problem

Given: Observed colors $C_{\mathbf{Observed}}$ from an image

Find: Illuminant $\mathbf{L}_{\text{illuminant}}$

Goal: Recover true surface colors $\mathbf{R}_{\text{surface}}$ by removing illuminant influence

Challenge: This is an ill-posed problem—infinite combinations of **R** and **L** can produce the same $\mathbf{C}_{\text{observed}}$

RELATED WORK

Method	Approach	Description	Limitations
Gray-World	Global mean assumption	• Computes average RGB across entire image • Assumes scene average = neutral gray • Simple illuminant estimation: L = (R̄, Ḡ, B̄)	**Fails with dominant colored objects** • Kitchen with red cabinets → incorrect red illuminant • No semantic understanding • Ignores object identity
White-Patch	Peak channel	• Finds brightest pixel in image • Assumes it's white under neutral light • Uses max(R, G, B) as illuminant estimate	**Highly sensitive to outliers** • Specular highlights mislead algorithm • Fails if brightest object is colored • Unreliable in low-light scenes
Bayesian (Gehler et al. 2008)	Probabilistic prior	• Learns illuminant distribution from training data • Uses Bayesian inference for estimation • Incorporates prior knowledge of natural lighting	**Requires large training data** • Dataset bias issues • Limited to seen lighting conditions • Slow inference time
Semantic CC (Afifi & Brown 2018)	CNN-based Black-box,	• End-to-end deep neural network • Direct image → illuminant mapping • Learns from millions of examples • No explicit physical modeling	**Data hungry & Computationally intensive** • Needs 10,000+ labeled images • Large model size (millions of parameters) • No interpretability • Poor generalization to new scenes

Dataset Creation and Validation

Dataset Filtering

Source: COCO train2017 dataset (118K images)
Criteria: Indoor scenes containing 2+ target object types

- Target objects: refrigerator, microwave, oven, sink, toilet

Quality filters: Remove low-quality, outdoor, and single-object images
Result: 5,644 diverse indoor scenes selected

Dataset Creation

Four illuminant types applied:
Red [2.2, 1.0, 0.7],
Blue [0.6, 1.0, 1.8],
Orange [2.0, 1.0, 0.6],
Yellow [1.7, 1.0, 0.8]

- Each normalized with green channel = 1.0

Advantage: Known ground truth illumination enables quantitative evaluation

BUILDING OBJECT PRIORS

Segment Object in Training Images:

- Use YOLO v8-seg to detect and mask each object instance.
- Apply mask refine (closing and erosion) to get filter high quality RGB pixels
- Remove dark/shadow pixels using brightness threshold

Quantize RGB Space into Voxels (8x8x8):

- Divide each channel into 8 bins i.e., bin size = 32
- Each pixel is mapped to the voxel as:

$$r = [R/32],\ g = [G/32],\ b = [B/32]$$

- Each voxel stores pixel count $n_c(v)$ and sum of RGB colors $s_c(v)$

Compute Voxel Centroids & Weights:

$$\text{Centroid } (v) = s_c(v)/n_c(v)$$

$$\text{Weight } (v) = n_c(v)/N_c$$

- where N_c is the total number of pixels for each class.

Build Semantix Prior:

Each class gets:

1. Set of representative color modes (voxel centroids)
2. Weights indicating stability/frequency
3. Global median color
4. Stored in an YAML file

METHODOLOGY

Diversity-Aware Dataset Creation

Problem: Most color constancy datasets lack indoor diversity

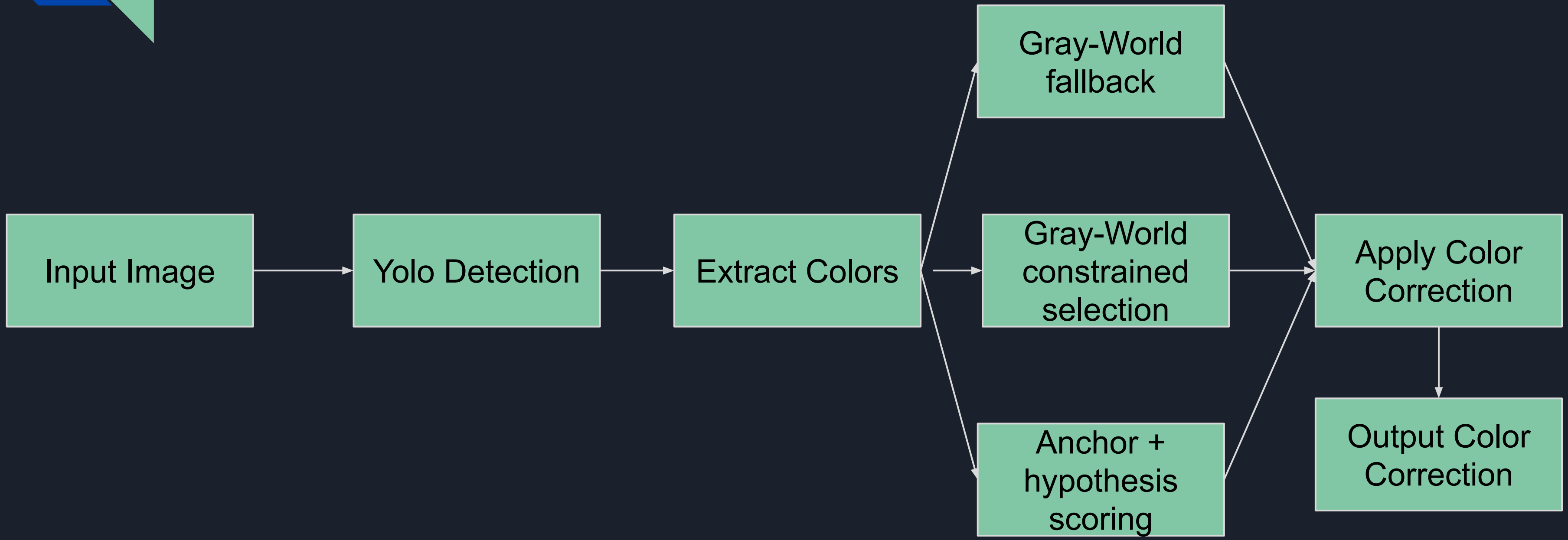
Semantic Prior Learning from Color Distributions

Goal: Model the typical color modes of each class by averaging pixel colors within 8x8x8 voxels.
Why: Objects have several color clusters, so a voxelized distribution captures these modes better than a single mean.
Example: For refrigerators, populated voxels often produce centroids corresponding to soft whites and light grays, reflecting their common true colors.

Adaptive Hypothesis Scoring

Challenge: How to estimate illuminant from detected objects?

MAIN PIPELINE



Multi Hypothesis Scoring

Goal: pick the illuminant L that makes most objects in the scene look consistent with their semantic color priors.

1. Anchor selection

- Choose a high confidence "anchor" object (largest mask with good priors).
- Use its observed mean color $C_{\mathbf{Obs,anchor}}$ and its prior modes

2. Generate candidate illuminants

- For each prior mode j of the anchor (voxel centroid C prior,j): $$\mathbf{L}_j = \frac{\mathbf{C}_{\text{obs,anchor}}}{\mathbf{C}_{\text{prior},j} + \epsilon}$$
- This says: "What illuminant would make the anchor look like its typical true color?"

3. Score each hypothesis using other objects
- For every other object iii, correct its mean color under $C_{\mathbf{j}}$: $$\mathbf{C}_{\text{corr},i,j} = \frac{\mathbf{C}_{\text{obs},i}}{\mathbf{L}_j + \epsilon}$$
- Check if $C_{\mathbf{corr,i,j}}$ falls inside that class's prior voxels (with a small RGB margin).
- If it fits, add its pixel count $N_{\mathbf{i}}$ to the score: $$S_j = \sum_{i \neq \text{anchor}} N_i \cdot 1_{\text{inside}}\left(\mathbf{C}_{\text{corr},i,j}, \mathcal{P}_i\right)$$

4. Final illuminant choice:
- Pick the Final Illuminant with highest score.

Results

2 Synthetic Tint Generated Images + 4 Real Images were added for validation sake

Results Qualitative

Original | Tint Image | Corrected Image

Color Constancy Image | Corrected Image

Thank you

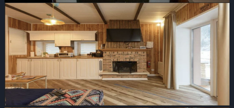
Questions?

Multi Illumination Smooth vs Many points of Illuminants (Limitations)

Original Image

Original Output from One illuminant

Results Quantitative

	MAE	MSE	PSNR
	12.326515197753906	329.1377868652344	22.95702616267224
	14.64657211303711	346.6778564453125	22.73154258267569
	17.39215850830078	868.1218872070312	18.74499655017302
	34.17348098754883	2210.340087890625	14.686212605171516
	22.557239532470703	1074.69873046875	17.81793625018494
	7.009693622589111	160.88584899902344	26.065625142325707