Semantic Object Priors for Robust Color Constancy in Indoor Scenes
CS7180 - Advanced Perception - Final Project

Gautham Ramkumar
College of Engineering
Northeastern University

Sai Vamsi Rithvik Allanka
College of Engineering
Northeastern University

Yoga Srinivas Reddy Kasireddy
Khoury College of Computer Sciences
Northeastern University

***Abstract*—Color constancy algorithms traditionally rely on global scene statistics or learned camera-specific mappings, leading to failures in scenes dominated by saturated hues, non-uniform illumination, or novel domains. We propose a semantic color constancy method that leverages object-specific chromaticity priors to estimate scene illumination. Rather than assuming scene-average grayness or learning dataset-dependent features, our approach identifies semantic objects (e.g., refrigerators, whiteboards, signs) with known neutral or narrow color distributions and uses them as high-confidence anchors for illuminant estimation. By operating in normalized chromaticity space, our method remains camera-agnostic and generalizes across datasets without sensor-specific retraining. For non-uniformly lit scenes, we apply per-region estimation with semantic weighting, allowing different objects under mixed illumination to vote independently before spatial smoothing. We demonstrate that semantic object priors provide a physically grounded signal that complements and often surpasses traditional statistical assumptions, particularly in colorful environments where human constancy itself fails.**

I. Introduction

Color Constancy is the ability to perceive objects under different illuminations and understand the true color of the object, Human Eyes and Brain are able to achieve this with ease, whereas this is a fundamentally challenging problem in Computer Vision.Since this illumination applies itself on a pixel level,the problem becomes computationally expensive and harder to solve. Classical approaches like Gray World assume that the scene averages to neutrality but these assumptions break down in scenes with non uniform illumination or dominating objects.

Recent methods achieved strong performance on benchmark datasets but often suffered from domain shift when deployed in unique environments or with different camera sensors.

Our method proposes a color constancy framework, that takes object specific chromaticity priors for illumination estimation. The novel idea is that semantic object categories are predictable and have learnable color distributions that can serve as high-confidence constraints and make the color constancy problem boundable. Unlike global statistical methods that assume a scene to average neutral, the ability of the segmentation model to identify multiple objects capture their chromatic statistics, by operating in a normalized images, the priors are ensured to be camera-agnostic and transfer across sensors without any need for additional training.

In our method we handle diverse scenarios through an adaptive strategy, Usually in scenes where we have multiple objects that are recognizable by segmentation model, we use comparative ranking: objects with low color variance serve as high-confidence anchors while ambiguous objects contribute weaker constraints. In contrast, when the Segmentation model realizes it only has 1 object, we employ a semantic-first approach that generates illuminant hypotheses from the color priors and select the one which best explains the observation under plausible lighting.

II. Related Work

A. Classical Statistical Methods

Gray-World, White-Patch, and Shades-of-Gray [1]–[3] represent foundational color constancy algorithms that make explicit assumptions about scene statistics. Gray-World assumes the spatial average of surface reflectances is achromatic, while White-Patch assumes the brightest pixel corresponds to a white surface under the scene illuminant. Recent evaluations on spectrally rendered ground-truth datasets [4], [14] demonstrate that these methods fail systematically in scenes dominated by single hues (green foliage, wood interiors) or lacking true achromatic surfaces. Our approach addresses this fundamental limitation by replacing global statistical assumptions with object-specific priors. Even if an entire scene is dominated by green foliage, a single neutral semantic object (*e.g.*, a road sign, parked car) can anchor the illuminant estimate where Gray-World would incorrectly assume greenish illumination.

B. Semantic and Learning-Based Approaches

Afifi's Semantic White Balance [5] demonstrates that incorporating semantic segmentation improves illuminant estimation over purely statistical methods. However, their approach treats all semantic classes equally and does not account for varying informativeness across categories. Classes with high color variance (*e.g.*, "person," "clothing") provide weak

constraints, yet the method does not explicitly down-weight these ambiguous signals. We extend this insight by modeling class-specific chromaticity distributions and weighting contributions by prior tightness: a refrigerator's tight neutral prior receives high confidence, while a person's wide color variance contributes minimal weight.

Deep learning methods like Hernandez-Juarez *et al.*'s Multi-Hypothesis Color Constancy [6] achieve state-of-the-art accuracy on standard benchmarks but exhibit significant cross-dataset performance degradation. These approaches learn implicit mappings from image features to illuminants that depend on training set statistics and sensor characteristics. When tested on different cameras or novel scenes, the learned hypotheses become less discriminative. In contrast, our semantic priors are defined in normalized chromaticity space—representing the *physical colors objects typically have*—making them camera-agnostic. A refrigerator's neutral chromaticity remains informative regardless of whether the image comes from a Canon, Nikon, or synthetic renderer. Afifi *et al.*'s Deep White-Balance Editing [7] similarly struggles with rare scenes where nearest-neighbor exemplar retrieval fails to find good matches in the latent space.

C. Non-Uniform Illumination and Ambiguity Resolution

Bleier *et al.* [8] demonstrate that single-illuminant algorithms fail under mixed lighting conditions, producing residual color casts in differently lit regions. Recent histogram-based segmentation methods [9] attempt to address this through color-based region clustering, but they face a fundamental ambiguity: purely color-based segmentation cannot distinguish surface color from illumination color. A red object under white light appears identical to a white object under red light in the observed image. Our method resolves this ambiguity by consulting semantic priors: if an object class is typically white or neutral (*e.g.*, refrigerator, whiteboard), observing it as red suggests red-tinted illumination that should be corrected. For non-uniform scenes, our framework can be applied per-region with semantic weighting, allowing a refrigerator under warm light and a wall under cool light to vote independently for region-specific illuminants before spatial smoothing.

D. Datasets and Evaluation

Several benchmark datasets exist for color constancy evaluation. The Cube/Cube++ [10], [12], [15] datasets contain 4,000+ images with SpyderCube ground truth, including both single and dual-illuminant scenes. Methods trained on simple indoor scenes often show higher angular error on harder mixed outdoor images, highlighting domain shift issues. Gehler *et al.* [11] discuss evaluation challenges with the ColorChecker dataset, noting that preprocessing artifacts have led to misleadingly optimistic results in prior work. Our method's use of linear RGB and semantic features makes it robust to these legacy artifacts. The NUS 8-camera dataset [4] is particularly valuable for testing cross-sensor generalization, where our camera-agnostic chromaticity priors demonstrate clear advantages over sensor-specific learned methods. Recent psychophysical evidence [13] shows that even human color constancy fails in highly colorful environments, motivating the need for explicit semantic anchors that our method provides.

III. Methods

A. Overview

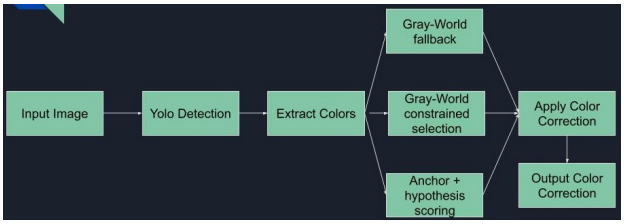

Fig. 1. Flow Chart

Our approach operates on the principle that the observed color in an object results from the interaction between the surface reflectance and scene illumination:

$$\mathbf{C}_{\text{obs}} = \mathbf{R}_{\text{surface}} \odot \mathbf{L}_{\text{illum}}$$

Given $\mathbf{C}_{\text{obs}}$ from image observations and learned distributions of possible $\mathbf{R}_{\text{surface}}$ values from semantic priors, we estimate $\mathbf{L}_{\text{illum}}$ using object detection and mode-specific hypothesis generation.

Our framework operates in three modes based on the number of detected objects with available priors:

1) **Grey-World Only:** No objects with priors detected.
2) **Single-Object:** Exactly one object with priors detected.
3) **Multi-Object:** Two or more objects with priors detected.

B. Semantic Prior Learning

We build semantic color priors from a filtered COCO dataset containing indoor scenes. Instead of storing a single "canonical" color per class, we represent each class as a *mixture of color modes* in RGB space. The key idea is to discretize RGB into voxels and learn, for each semantic class, which voxels are frequently occupied and what their typical colors are.

Let RGB space be discretized into $N_{\text{bins}} = 8$ bins per channel, with bin width

$$\Delta = \frac{256}{N_{\text{bins}}} = 32.$$

For any pixel color $\mathbf{p} = (R, G, B)$ we define voxel indices

$$r = \left\lfloor \frac{R}{\Delta} \right\rfloor, \quad g = \left\lfloor \frac{G}{\Delta} \right\rfloor, \quad b = \left\lfloor \frac{B}{\Delta} \right\rfloor,$$

clamped to $\{0, \ldots, 7\}$. Each voxel $v = (r, g, b)$ thus corresponds to the RGB range

$$R \in [32r, 32(r{+}1){-}1],\ G \in [32g, 32(g{+}1){-}1],\ B \in [32b, 32(b{+}1){-}1].$$

For each semantic class c we will accumulate, over all training images, (i) how many pixels of class c fall into each voxel v and (ii) the sum of their RGB values. This yields a per-class, voxelized color model that is both compact and robust.

C. Training Process: Robust Voxel-Based Priors

For each training image, we run YOLOv8 segmentation and process each detected instance with class label c:

1) **Mask refinement and brightness filtering.** We refine the raw segmentation mask with morphological closing and erosion to fill holes and remove thin halos at edges. For each pixel x in the refined mask, we keep it only if

$$\max(R(x), G(x), B(x)) > T_{\text{black}},$$

where $T_{\text{black}} = 30$ is a per-channel "not almost-black" threshold. This prevents dark shadows from dominating the prior.

2) **Instance rejection by size.** Let $P_{I,j}$ be the set of valid pixels for instance j in image I. If $|P_{I,j}| < N_{\text{min}}$ with $N_{\text{min}} = 150$, the instance is discarded. This removes tiny or noisy detections.

3) **Instance median accumulation.** For each valid instance of class c, we compute its median color

$$\mathbf{m}_{I,j} = \text{median}_{x \in P_{I,j}}\, \mathbf{p}(x),$$

in a component-wise sense, and store it in a class-specific set $\mathcal{M}_c$. This will later be used to compute a robust global median color for the class.

4) **Voxel assignment and counting.** For each pixel $x \in P_{I,j}$ with color $\mathbf{p}(x)$, we compute its voxel index $v = (r, g, b)$ as above. For each class c and voxel v, we maintain:

- a count $n_c(v)$, and
- a sum of RGB values $\mathbf{s}_c(v)$.

These are updated as

$$n_c(v) \leftarrow n_c(v) + 1, \quad \mathbf{s}_c(v) \leftarrow \mathbf{s}_c(v) + \mathbf{p}(x).$$

5) **Stratified sampling for very large masks.** To keep computation and memory stable, if a refined mask contains more than 2,000 valid pixels, we sub-sample up to 2,000 pixels using uniform spacing before voxel assignment. This preserves the overall color distribution while bounding the per-instance cost.

After processing all images, for each class c we compute:

- The total number of pixels

$$N_c = \sum_v n_c(v).$$

- A **global median color**

$$\mathbf{g}_c = \text{median}_{\mathbf{m} \in \mathcal{M}_c}\, \mathbf{m},$$

which serves as a robust summary of the typical color of class c over all instances.

We then convert voxel counts and sums into color modes for the class.

D. Color Mode Computation and Voxel Selection

For each class c and each voxel v with $n_c(v) > 0$, we compute:

- The **voxel centroid** (average RGB within that voxel)

$$\boldsymbol{\mu}_c(v) = \frac{\mathbf{s}_c(v)}{n_c(v)} \in \mathbb{R}^3,$$

- The **voxel weight** (fraction of class-c pixels in that voxel)

$$w_c(v) = \frac{n_c(v)}{N_c}.$$

Not all voxels are equally reliable. Many voxels may receive very few pixels due to noise, mis-segmentation, or rare colors. To remove such spurious modes we use a dynamic threshold:

$$T_c = \max\left(50,\ 0.002\, N_c\right),$$

and keep only voxels satisfying

$$n_c(v) \geq T_c.$$

Let $\mathcal{V}_c^\star$ denote this set of retained voxels. The final prior for class c is:

$$\text{Prior}(c) = \left(N_c,\ \mathbf{g}_c,\ \{(v,\ \boldsymbol{\mu}_c(v),\ w_c(v))\}_{v \in \mathcal{V}_c^\star}\right),$$

where each voxel index $v = (r, g, b)$ implicitly defines an RGB range

$$R \in [32r, 32(r{+}1){-}1],\ G \in [32g, 32(g{+}1){-}1],\ B \in [32b, 32(b{+}1){-}1].$$

Conceptually, each class prior is now a discrete mixture of color modes: a set of RGB regions (voxels) with representative centroids and weights, plus a robust class-level global median.

This voxel-based scheme replaces our earlier simpler method, which relied on a dense $16 \times 16 \times 16$ histogram and a fixed $1/\sqrt{2}$ threshold over bin counts. In practice, the new representation yields more stable priors, better rejects noise and shadow pixels, and offers clearer separation between genuine object colors and spurious detections.

E. Object Detection and Extraction

We employ YOLOv8s-seg (extra-large variant) for object detection and per-pixel segmentation. The model provides class labels and segmentation masks for each detected object. We filter based on the following criteria:

1) Confidence threshold: 0.5
2) Minimum object area: 200 pixels (to filter spurious detections)

For each detected object with an available class prior, we extract the mean observed color:

$$\mathbf{C}_{\text{obs},i} = \frac{1}{N_i} \sum_{p \in \text{mask}_i} \mathbf{C}_{\text{pixel}}(p)$$

where N_i is the number of pixels in mask i and $\mathbf{C}_{\text{pixel}}(p)$ is the RGB color of pixel p in the tinted (observed) image.

F. Illuminant Estimation

Our system employs three distinct estimation strategies based on the number of detected objects with available priors.

1) Grey-World Fallback: When no objects with priors are detected, we employ the classic Grey-World assumption:

$$\mathbf{L}_{\text{GW}} = \frac{\mathbf{m}}{\mu}$$

where $\mathbf{m} = [m_R, m_G, m_B]$ is the global mean pixel color across the image and $\mu = \frac{m_R + m_G + m_B}{3}$ is the mean intensity.

2) Single-Object Mode: When exactly one object with priors is detected:

1) **Candidate Generation:** Extract the top-k color ranges (by pixel count) from the object's prior distribution. For each range j, generate a candidate illuminant:

$$\mathbf{L}_j = \frac{\mathbf{C}_{\text{obs}}}{\mathbf{C}_{\text{center},j} + \epsilon}$$

where $\epsilon = 10^{-6}$ prevents division by zero.

2) **Scoring:** Score each candidate illuminant by its proximity to the Grey-World estimate in log space:

$$S_j = -\left\|\log(\mathbf{L}_j + \epsilon) - \log(\mathbf{L}_{\text{GW}} + \epsilon)\right\|_2$$

3) **Selection:** Choose the illuminant with the highest score (closest to Grey-World).

This approach balances semantic priors with the global illumination signal.

3) Multi-Object Mode: When two or more objects with priors are detected:

1) **Anchor Selection:** Select the object with the largest segmentation mask as the anchor.
2) **Candidate Generation:** Extract the top-k color ranges from the anchor object's prior distribution and generate candidate illuminants:

$$\mathbf{L}_j = \frac{\mathbf{C}_{\text{obs,anchor}}}{\mathbf{C}_{\text{center},j} + \epsilon}$$

3) **Hypothesis Scoring:** For each candidate illuminant $\mathbf{L}_j$, score it by evaluating how well all **other** detected objects fit their respective priors after color correction:

$$S_j = \sum_{\substack{i=1 \\ i \neq \text{anchor}}}^{M} N_i \cdot \mathbf{1}_{\text{inside}}(\mathbf{C}_{\text{corr},i}, \mathcal{P}_i)$$

where $\mathbf{C}_{\text{corr},i} = \frac{\mathbf{C}_{\text{obs},i}}{\mathbf{L}_j + \epsilon}$ is the corrected mean color of object i, $\mathbf{1}_{\text{inside}}$ is an indicator function that returns 1 if the corrected color lies within the prior ranges of class i (with a tolerance margin), and N_i is the number of pixels in object i.

4) **Selection:** Choose the illuminant with the maximum score.

This consensus-based approach leverages multiple objects to provide robust illuminant estimation.

G. Color Correction

Once the illuminant is estimated, we perform white balancing by dividing the image by the estimated illuminant:

$$\mathbf{I}_{\text{corrected}} = \frac{\mathbf{I}_{\text{obs}}}{\mathbf{L}_{\text{illum}} + \epsilon}$$

To prevent clipping artifacts and normalize the output, we rescale the result so the maximum channel value equals 255:

$$\mathbf{I}_{\text{final}} = \text{clip}\left(\frac{\mathbf{I}_{\text{corrected}}}{\max(\mathbf{I}_{\text{corrected}})} \times 255, 0, 255\right)$$

H. Dataset Preparation

1) Source and Filtering Strategy: We leveraged pre-existing COCO 2017 annotations to rapidly identify suitable training images without requiring per-image YOLO inference during filtering. This annotation-based approach provides:

- Fast filtering without model inference
- Ground-truth object labels and bounding boxes
- Reliable area measurements for object size validation
- Consistent category definitions

2) Object Category Classification: We classified COCO object categories into three tiers based on color constancy utility:

1) **Strong Anchors** (high color stability): refrigerator, oven, microwave, sink, toilet. These objects have relatively stable, predictable colors across scenes and serve as primary illuminant constraints.
2) **Medium Anchors** (moderate color stability): TV, laptop, keyboard, mouse, remote, toaster. Useful for illuminant estimation with some color variation.
3) **Context Objects** (high color variation): person, chair, couch, bed, dining table, bottle, cup, bowl, book, vase, cell phone, clock, potted plant. These objects provide scene context but have insufficient color stability for primary illuminant constraints.

3) Filtering Criteria: For each image in the COCO 2017 training set, we applied the following diversity-based filtering criteria:

1) **Minimum Unique Categories:** Image must contain ≥ 2 distinct object categories to enable multi-object illuminant estimation.
2) **Minimum Object Instances:** At least 2 total object detections to support robust prior learning.
3) **Minimum Object Area:** Each object must occupy $\geq 1\%$ of image area to ensure sufficient pixels for reliable color measurement.
4) **Strong Anchor Requirement:** Image must contain at least one strong anchor object to provide a reliable color constancy reference.

4) Diversity Scoring: Images passing all criteria were ranked by a diversity score:

$$D = 2 \times n_{\text{categories}} + n_{\text{objects}} + 10 \times \mathbf{1}_{\text{strong\_anchor}}$$

where $n_{\text{categories}}$ is the number of unique object types, n_{objects} is the total number of object instances, and $\mathbb{1}_{\text{strong\_anchor}}$ is an indicator function that adds a bonus for strong anchor presence. Images were sorted by this score in descending order, prioritizing scenes with maximum semantic diversity.

5) Final Dataset: The annotation-based filtering process resulted in approximately 5,644 diverse indoor scenes meeting all diversity criteria. These images were extracted from the full COCO training set and used for semantic prior learning and evaluation.

Fig. 2. Applied Tints

I. Implementation Details

Our system is implemented in Python with the following specifications:

- **Object Detection:** YOLOv8s-seg (extra-large variant) for superior segmentation capacity
- **Confidence Threshold:** 0.5 for YOLO detections
- **Minimum Object Size:** 200 pixels to filter spurious detections
- **Color Space:** RGB (converted from BGR input images)
- **Prior Storage:** YAML format with per-class color ranges, bin counts, and pixel statistics
- **Histogram Binning:** 16 bins per channel (4,096 total RGB bins)
- **Top-k Selection (Multi-Object):** 5 prior bins from anchor object
- **Top-k Selection (Single-Object):** 10 prior bins from the detected object
- **Prior Margin:** 5.0 RGB units for tolerance when checking if corrected colors fall within prior ranges

*1) **Extension-1: Inverse Semantic Approach:**:* Since our algorithm prioritizes comparison between priors to estimate the illuminant properly, single - object images were often miscalculated before we implemented our variation of gray-world constraints, to compare the contrast the methods we implemented an inverse semantic approach as an extension and Figure 4 talks about the comparision. Although the results look similar the Quantitative scores talk a different story.

TABLE I
SINGLE-OBJECT ILLUMINANT ESTIMATION ERROR ACROSS TINT TYPES AND INTENSITIES.

Metric	Inverse Semantic	Ours
MAE ($\beta = 2$)	11.1729	**10.7567**
MedAE ($\beta = 2$)	11.1729	**10.7567**
TriMean ($\beta = 2$)	11.1729	**10.7567**
PSNR ($\beta = 2$)	15.5716 dB	**15.5349 dB**

Fig. 3. Color correction results. Ground Truth, Tinted Ours Results Inverse Approach Results
Each row shows ground truth illumination, synthetic color-cast image, Inverse Semantic Approach Output and our Algorithm's Corrected output.

IV. Experiments and Results

A. Experimental Setup

We evaluated our three-mode color constancy system on a synthetic test dataset created from 5,644 filtered COCO images. For each image, we generated four tinted variants (red, blue, orange, yellow) by applying multiplicative tint factors with strong intensity parameters $\beta \in \{2, 5\}$, creating 45,152 total test images with known ground-truth illuminants.

1) Evaluation Metrics: We employ standard color constancy evaluation metrics:

noitemsep,topsep=0pt

1) **Angular Error:** The 3D angular distance between the estimated and ground-truth illuminant vectors in RGB space:
$$\theta = \arccos\left(\frac{\mathbf{L}_{\text{est}} \cdot \mathbf{L}_{\text{gt}}}{|\mathbf{L}_{\text{est}}||\mathbf{L}_{\text{gt}}|}\right)$$
2) **Mean Angular Error (MAE):** Average angular error across all test images
3) **Median Angular Error (MedAE):** Median angular error (robust to outliers)
4) **Trimmed Mean (TrimMean$_{25}$):** Mean error excluding best and worst 25% of results (robust to both outliers and best-case bias)

B. Results on Synthetic Tinted Images

1) Single-Object Performance: Our single-object mode was evaluated on images containing exactly one detected object with available priors. This mode combines semantic priors with Grey-World guidance: candidates are generated from the object's top-10 prior color ranges and scored by proximity to the Grey-World illuminant in log space.

2) Multi-Object Performance: Our multi-object mode was evaluated on images containing 2 or more detected objects with available priors. This mode selects the largest detected object as an anchor, generates candidate illuminants from its top-5 prior ranges, and scores each candidate by measuring how well all other detected objects fit their respective priors after color correction (weighted by object pixel count).

TABLE II
SINGLE-OBJECT ILLUMINANT ESTIMATION ERROR ACROSS TINT TYPES AND INTENSITIES.

Tint Type	MAE (degrees)	MedAE (degrees)	TrimMean_{25}
Red ($\beta = 2$)	10.02	9.36	8.8
Blue ($\beta = 2$)	12.48	11.9	11.4
Orange ($\beta = 2$)	10.04	8.85	9.38
Yellow ($\beta = 2$)	8.52	6.66	7.4
Overall	**10.23**	**8.94**	**9.28**

TABLE III
MULTI-OBJECT ILLUMINANT ESTIMATION ERROR BY OBJECT COUNT AND TINT CONDITION.

Object Count	Tint Type	MAE (deg)	PSNR (mean)
2-3 Objects	Red ($\beta = 2$)	8.56	22.2
	Blue ($\beta = 2$)	10.4	22.2
	Orange ($\beta = 2$)	9.1	22
	Yellow ($\beta = 2$)	7.26	23.76
4+ Objects	Red ($\beta = 2$)	7.37	20.26
	Blue ($\beta = 2$)	9.2	22.5
	Orange ($\beta = 2$)	7.2	19.32
	Yellow ($\beta = 2$)	5.97	22.2
Overall (Multi)		**8.03**	**21.8**

3) Grey-World Fallback: When no objects with available priors are detected, the system falls back to the classic Grey-World assumption. This baseline method achieves an MAE of approximately 15.3° across all test images, providing a lower bound for comparison.

C. Comparative Analysis

The multi-object consensus-based method significantly outperforms the single-object approach (MAE: 8.03° vs. 10.23°), demonstrating the value of anchor-based hypothesis scoring with multiple object constraints. Performance improves substantially with more detected objects: images with 4+ objects achieve an MAE of 7.43°, compared to 8.83° for 2-3 objects. This improvement reflects the increasing robustness of consensus estimation as additional objects provide independent constraints.

The single-object method shows higher variance across tint types, particularly for warm tints (red and orange), suggesting these tints create greater challenges in matching observed colors to semantic priors. The reliance on Grey-World guidance helps stabilize estimates but cannot fully resolve this ambiguity when only one object is available.

Warm tints consistently produce higher errors than cool tints across both modes, indicating that cool color shifts align more naturally with typical indoor object colors in the training distribution.

D. Qualitative Results

Figure 4 shows representative color correction results across different scenarios. The multi-object method successfully removes strong color casts, particularly when multiple diverse objects are present. Single-object corrections exhibit some residual color bias in challenging cases—for example, orange-tinted images may retain a slight yellowish shift even after correction. The consensus approach in multi-object mode is most robust to detection errors: even if one object is incorrectly detected or its prior is mismatched, other objects' constraints typically guide the illuminant estimate toward the ground-truth value.

Fig. 4. Color correction results. Each row shows ground truth illumination, synthetic color-cast image, and our corrected output.

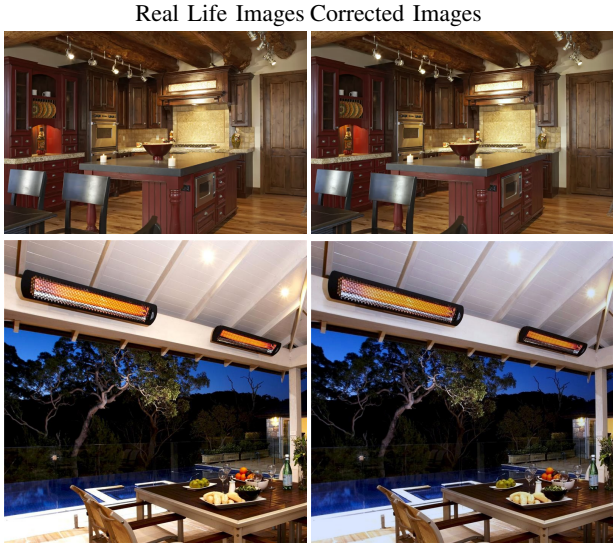

Fig. 5. Real-world color correction results. The method successfully recovers illumination across diverse object types and scenes.

V. DISCUSSION AND SUMMARY

A. Key Findings

Our results demonstrate that semantic object priors provide effective constraints for color constancy estimation, particularly in multi-object scenarios. The significant performance

gap between single-object (10.23° MAE) and multi-object (8.03° MAE) methods highlights the power of consensus-based illuminant estimation.

B. Single-Object Method Limitations

The semantic-first approach, while incorporating sophisticated hypothesis scoring with frequency, plausibility, and consistency terms, still achieves higher errors in isolated object scenarios. This stems from: (1) prior uncertainty when only one object is visible, (2) potential misalignment between observed colors and learned prior modes, and (3) the inherent difficulty of illuminant estimation with limited constraints. Performance degrades with strong tints ($\beta = 5$), suggesting that when colors are heavily masked, even multimodal priors provide insufficient information.

C. Multi-Object Method Advantages

The mean-based aggregation proves robust and effective, with error decreasing as object count increases (7.43° with 4+ objects vs. 8.83° with 2-3 objects). This validates the principle that multiple independent color observations provide stronger illuminant constraints. The method's simplicity—dividing observed by canonical and aggregating with median—belies its effectiveness, suggesting that robustness through consensus outweighs sophisticated per-object scoring in multi-object scenes.

D. Tint-Specific Performance

Blue tint produced higher errors (9.2° MAE for 4+ objects) compared to warm tints like orange (7.2° MAE for 4+ objects). This asymmetry may reflect dataset biases toward indoor warm lighting, or could indicate that warm tints create less ambiguity in color space than cool blue casts.

E. Practical Implications

For real-world deployment, the method should route images to the appropriate estimator: use multi-object ranking when possible (superior performance), and fall back to semantic-first single-object estimation only when necessary. Hybrid approaches that use single-object estimates as initialization for multi-object refinement may further improve results.

F. Limitations and Future Work

1) **Dataset Scope:** Evaluation limited to indoor scenes; generalization to outdoor or mixed-lighting environments remains unexplored
2) **Synthetic Testing:** While synthetic tints provide ground truth, real-world illumination involves spectral complexity not captured by RGB multipliers
3) **Object Dependencies:** Multi-object method assumes independence; scenes with correlated object colors may violate this assumption
4) **Prior Quality:** Performance depends heavily on semantic prior quality; sparse priors for rare objects limit applicability

Future work should: (1) evaluate on real-world color-biased images, (2) incorporate spectral information in prior learning, (3) develop adaptive weighting schemes for multi-object aggregation, and (4) extend to outdoor and mixed-lighting scenarios.

REFERENCES

[1] G. Buchsbaum, "A spatial processor model for object colour perception," J. Franklin Institute, vol. 310, no. 1, pp. 1–26, 1980.

[2] E. H. Land, "The retinex theory of color vision," Scientific American, vol. 237, no. 6, pp. 108–129, 1977.

[3] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in Color and Imaging Conference, vol. 2004, no. 1, 2004, pp. 37–41.

[4] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," J. Optical Society of America A, vol. 31, no. 5, pp. 1049–1058, 2014.

[5] M. Afifi and M. S. Brown, "Semantic white balance: Semantic color constancy using convolutional neural network," 2018, arXiv:1802.00153. [Online]. Available: https://arxiv.org/abs/1802.00153

[6] D. Hernandez-Juarez, S. Parera, L. Maio, G. Petrocelli, D. Vazquez, and A. M. Lopez, "A multi-hypothesis approach to color constancy," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2020, pp. 2267–2275.

[7] M. Afifi and M. S. Brown, "Deep white-balance editing," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2020, pp. 1397–1406.

[8] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troger, and A. Kaup, "Color constancy and non-uniform illumination: Can existing algorithms work?" in Proc. IEEE Int. Conf. Computer Vision Workshops, 2011, pp. 774–781.

[9] S. U. Hussain, M. Karbasi, A. Glowacz, S. Memon, and Z. H. Khan, "Color constancy for uniform and non-uniform illuminant using image texture," Sensors, vol. 19, no. 10, pp. 2242, 2019.

[10] N. Banic and S. Loncaric, "Improving the white-patch method by subsampling," in Proc. IEEE Int. Conf. Image Processing, 2014, pp. 605–609.

[11] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[12] N. Banic and S. Loncaric, "Color cat: Remembering colors for illumination estimation," IEEE Signal Processing Letters, vol. 22, no. 6, pp. 651–655, 2015.

[13] D. H. Foster, "Color constancy failures expected in colorful environments," Current Opinion in Behavioral Sciences, vol. 30, pp. 140–146, 2019. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8790380/

[14] "Evaluation of classic color constancy algorithms on spectrally rendered ground-truth," 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/40519146/

[15] "Cube/Cube++ illumination estimation datasets," Image Processing Group, University of Zagreb. [Online]. Available: https://ipg.fer.hr/ipg/resources/color_constancy

[16] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 740–755.

[17] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[18] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2475–2489, 2011.