Preprints.org

A Comprehensive Review of Deep Learning based Illumination Estimation

Swarna Sethu , Jayanthi Devaraj , Dongyi Wang [*]

Review

A Comprehensive Review of Deep Learning Based Illumination Estimation

Swarna Sethu [1], Jayanthi Devaraj [2] and Dongyi Wang [1,3,*]

[1] Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, 72701, AR, USA

[2] Department of Information Technology, Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, 602105, Tami Nadu, India

[3] Department of Food Science, University of Arkansas, Fayetteville, 72701, AR, USA

* Correspondence: dongyiw@uark.edu; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials)

Abstract: Deep learning (DL) models have been recently widely used to extract task-oriented patterns from large scale of datasets, and to improve the data image understanding and analysis accuracy in many different decision-making processes for tasks such as image classification, segmentation, detection, and so on. However, in practice, the performances of DL models are easily affected by environmental illumination conditions. Conversely, DL models can also be utilized for extracting the illumination hints from the images, and these hints are critically useful for improving the model robustness, classifying the environmental scenes, estimating scene depth information, and rendering 3D objects. In this study, an extensive and exhaustive review is carried out for DL based color constancy, indoor and outdoor illumination estimation, and image depth estimations with the considerations of strengths and weaknesses of DL models. This study also explores the different network designs and the paradoxes in parameter optimization during the model training. Current technology barriers involved in implementing these models and recommendations to overcome these barriers are also suggested in the review.

Keywords: data preprocessing; feature extraction; deep learning; illumination estimation; color constancy

1. Introduction

Color constancy is taken as the effect whereby the perceived or apparent color of an object surface remains constant despite changes in the intensity and spectral composition of the illumination (Foster, 2011). Human observers have the capabilities to discriminate and to remove the effects of changes in scene lighting and scene contents (Maloney and Brainard, 2010). However, achieving color constancy for machine vision system and machine learning models are critically challenging especially in natural conditions, which could affect the performance of deep learning models in basic classification and object detection tasks (Afifi et al., 2019). An important computational color constancy task is the illumination estimation for illumination parameter, such as the radiances and reflectance parameters, whose result can be used for photometric transformations in the image preprocessing step for a variety of real-world computer vision applications such as virtual reality, entertainment, digital museums, post-production of films, video gaming design, graphic design etc. Illuminant color estimation also plays a major role in effecting the performance of object tracking and object recognition tasks (Lou et al., 2015). This paper conducts a comprehensive review of the role of illumination estimation in the computer vision domain, and the current advances deep learning-based illumination estimation. In the following, Section 2 introduces the review methodology and scope and the motivation of the paper. Section 3 summarizes the current literature in illumination estimation methods, especially deep learning methods, and their applications. Section 4 listed some common challenges in adopting the illumination estimation models in practice, and some potential solutions are offered to overcome the barriers.

2. Review Methodology and Background

2.1. Review Scope and Strategies

Multiple sources of data are used for extracting the valuable content of literature study based on a keyword search strategy. To be specific, the articles are retrieved by using search items in various databases or search engines such as IEEE, science direct, Springer, google scholar, Scopus etc. Figure 1 shows the detailed methodologies followed in the literature study.
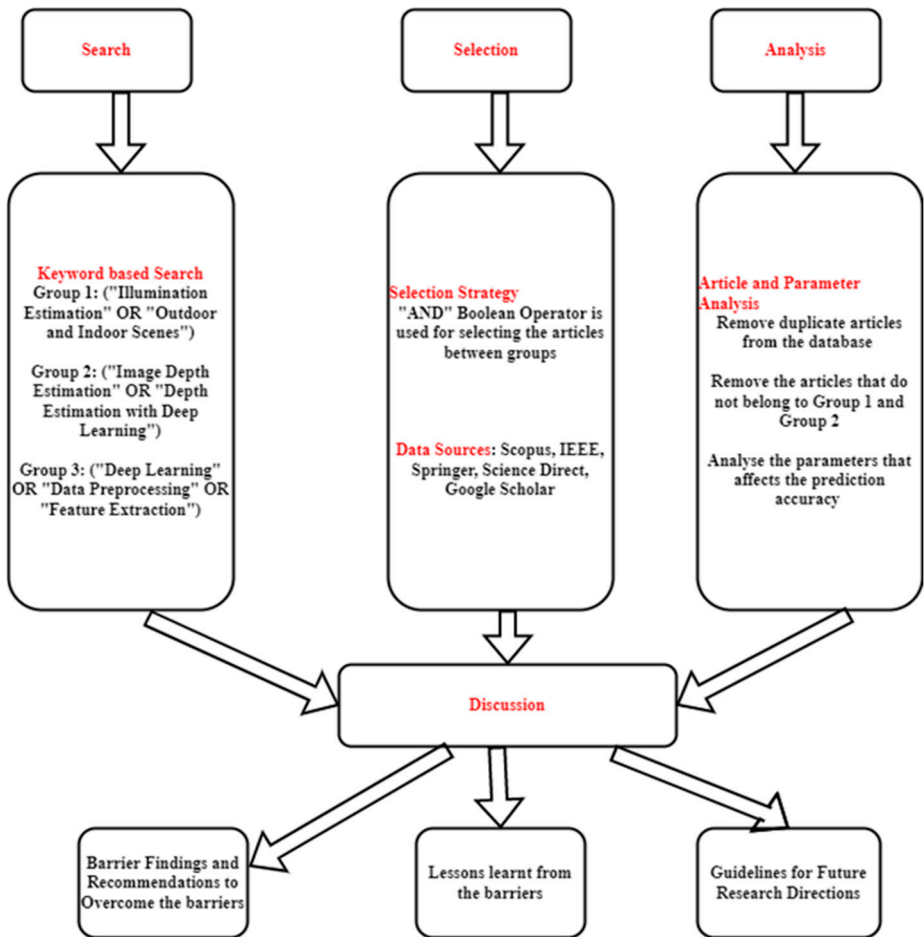

Figure 1. Framework for Literature Review.

There are three phases of search procedure for performing the extensive literature study: Phase 1 focuses on the accurate and efficient indoor and outdoor scenes estimation, and color constancy techniques for illumination estimation. Phase 2 searches different computational models for estimating the depth information from single RGB images based on illumination hints. Phase 3 specifically searches the role of deep learning models played in illumination and depth estimations along with corresponding pre-processing and feature extraction techniques. Based on the literature search, the barriers in the existing literature models are identified, and the future prospects and recommendations to overcome these challenges are formulated.

2.2. Research Gap and Motivations

It is well-accepted that DL models outperform traditional image processing and machine learning models in handling huge volumes of image or video data effectively (Sarker, 2021). Specifically, DL models has the capabilities to estimate the illumination parameters on different indoor or outdoor scenes, and there are a variety of related datasets available including images in different color spaces or with different color content under indoor or outdoor environments for

illumination estimation studies. This review firstly summarizes the recent advances for illumination and scene estimation models and discusses the deep learning-based color constancy techniques along with some other statistical techniques and optimization algorithms. The novelty of the extensive review involves: 1. Make a comprehensive and extensive review of illumination estimation, color constancy, indoor and outdoor scene estimation, and image depth estimation from single RGB image. 2. Summarize the roles of models, preprocessing techniques and feature extraction techniques played in improving the model performances. 3. Summarize different error metrics to evaluate the performances of deep learning models. 4. Identify barriers in the existing literature and make recommendations to potentially overcome these barriers.

2.3. Challenges in Color Constancy

However, color constancy and illumination face many challenges in real world environments. First, color constancy likely fails on some surfaces or parts of surfaces in natural outdoor environments considering the surface material properties (Foster and Reeves, 2022). Both estimated color differences and deviations in spatial ratios of cone excitations are correlated with each other, which is independent of the observer's adaptation. With any individual scene, either outdoors or in the laboratory, color differences detection also depends on an observer's color awareness, the observers' view scenes and on how global color properties guide attention (Foster and Reeves, 2022).

There are some other factors can affect color constancy performance, including the changes in natural illumination; the spatial distributions of surfaces within scenes; and the role of cognitive discounting. Natural illumination changes in many forms, and some changes could be sequential and rapid, as with a cloud passing over the sun or a shift in shadows cast more locally; some changes can be spatial, with light shining objects from different directions, or variegated with surfaces partly in direct sunlight and partly in shade. Thresholds for detecting color differences in these conditions are hard to be determined. On contrary, when illumination changes occur over minutes or hours, as with the spectrum of the direct beam during solar elevation, thresholds depend much more on experiments or human experience. The spatial illumination differences can result in surface color differences depending on the illumination scattered over the scene or concentrated in one area. Since localized reflected light changes can attract observers' attention by processes that are largely automatic and unconscious. Surface area may not be decisive, as long as it can be resolved (Foster et al., 2001). Similar considerations need to be applied to the effect of changes in color relations.

In practice, even detectable color differences can be discounted by higher-level cognitive processes. Experimentally, with suitable instructions, observers can indeed separate judgements of color appearance from judgements of the properties of objective reflecting surfaces under different illuminants. Color constancy may then be construed as the capacity to associate colors with particular surface spectral reflectance's, in so far as they are defined (Foster, 2011). Crucially, the capacity to make judgements about the origin of perceived changes seems possible only because differences in appearance are attributable to illumination changes that are not coextensive with object surfaces.

3. Deep Learning Methods and Applications in Illumination Estimation

Deep learning has played a major role in variety of computer vision tasks, and the DL models can automatically perform feature extractions and generate complex high level feature representations for images and signals. Specifically, multiple papers have been published recently which report the deep learning models can accurately estimate the illumination parameters from digital images, and the extracted illumination information can benefit many downstream applications. For instance, (Jin et al., 2021) demonstrated that in augmented reality applications, harmonic coefficients can be inferred in real-time utilizing deep learning models, and the coefficients can estimate the light settings (multiple light sources with distinct light intensities) of the real-world environment (Marques et al., 2018). The illumination hint can also be extracted from the depth information, and vice versa. Figure 2 shows an example illumination-based depth estimation diagram. Different illuminants and illuminant parameters can be estimated from input images, and the illumination settings can be used for correct image color. Based on the differences of ground truth

color constancy images and the corrected images, the depth information can be estimated from color images. In contrast, with the additional depth information, the illumination estimation results can be improved. The followings of this section will review literatures for computational methods and applications for illumination, scene, and depth estimations in indoor and outdoor settings.
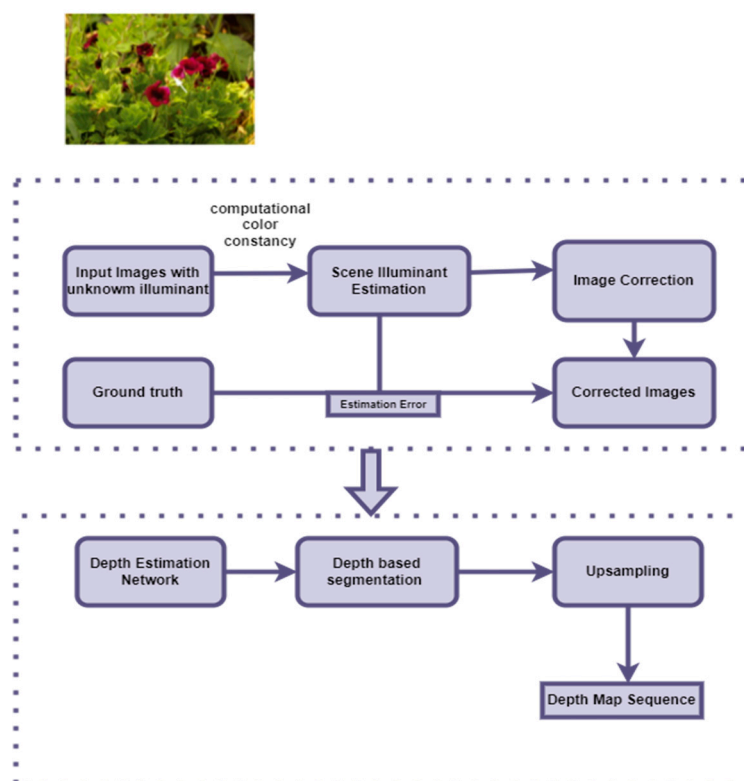
Figure 2. Color constancy for multiple illuminants estimation and depth map generation.

3.1. Color Constancy Methods

As discussed, computational color constancy is a prerequisite for many computer vision applications. Different light sources affect the appearance and color of the object. For human eyes, light sources can be identified with extreme scene light distributions, but the task is extremely challenging for the intelligent computer systems. By estimating illumination settings, computational color constancy can be achieved by transforming the original images utilizing the illumination parameters. Generally, the illumination estimation algorithms can be categorized as physical modelling methods, statistical methods, and learning based methods.

The physics-based methods mainly based on modelling the physical interactions between the light and the surfaces. There are two interaction models widely used, with the first one considering only diffuse reflection (Lambertian reflectance) and the second dichromatic model considering both diffuse reflection and specular reflections. However, the physical modelling may only be suitable for ideal conditions, and in practice, the modelling will be easily affected by noises, and it is very challenging to create the dichromatic lines for each surface (Tan et al., 2004). Statistical based methods can avoid the comprehensive modelling of surface and light interactions, and utilizes the empirical assumptions to describe the color distributions under common lights. There are three common statistical based methods, including gray-world, white-patch and gray-edge algorithms. Gray world assumption is based on that the average reflectance in the scene is gray color and the white-patch assumption is based on that the maximum reflectance is the perfect white reflector. Gray-edge assumption is an extension of gray world assumption, which assumes the average of the reflectance differences in a scene is achromatic. Gray-edge assumption is claimed to be performs better than gray world assumption considering local information (Van et al., 2007). Learning based color constancy and illumination estimation require to learn the statistical models from training data

(S. Bianco et al., 2015), this distinction partially determines the application scope of the algorithms real-world scenarios. For Gamut mapping method, a classic learning method, which in a preliminary phase, chooses a canonical illuminant and collect as many object surfaces as possible under the given illuminant (Forsyth, 1990). Then, given an input image with an unknown illuminant, the illuminant can be estimated from the canonical gamut and produces the most colorful and realistic scenes. Traditional machine learning models can also be trained to utilized low level image feature to predict the illumination parameters (P. V. Gehler et al., 2008), but CNN models and their variants, such as CNN based regression (Lou 2015), bias correction and fast Fourier integrated model can provide better prediction accuracy (Afifi 2021). These learning-based methods can also be fused with physical methods for better prediction performance (Schaefer et al., 2005).

3.2. Illumination Estimation Applications

Illumination estimation has become a critical task for building robust computer vision understanding models in real world applications and combining virtual objects with the real scenes in the simulation environments (Jin et al., 2021). The appearance of the scenes like brightness, color and shadows are determined by the illumination settings. In the indoor environment, the simplest illumination representation can be the spherical harmonic function which uses a small number of coefficients to represent the surrounding complex ambient light, and these coefficients can be used to reconstruct the illumination environment when rendering. The calculation of the spherical harmonic coefficient is based on the idea of probability theory, that is, the use of "finite" to estimate "infinite'. (Jin et al., 2021) proposes a real-time rendering technology, belonging to the category of precomputed radiative transfer (PRT). PRT is one type of computer graphics technique used to render a scene in real time with complex light interactions being precomputed to save time widely used in the field of game graphics rendering to simulate complex scene lighting quickly and in real time. Different illumination conditions may also affect the image quality. High dynamic range (HDR) indoor environment setup utilizes multiple frames from different illumination conditions to improve the dynamic range of the image. DL based illumination estimation can assist to predict high dynamic range images from low dynamic range (LDR) images. (Gardner et al., 2017) utilizes convolutional neural networks (CNN) to predict the light from a single image, where the CNN model is trained in two stages. First, light sources are detected from a LDR panoramas database to generate target light masks (J. Xiao et al., 2012). Second, the network can be fine-tuned on a novel HDR panorama dataset, which can be used for predicting light intensities. Compared to conventional machine learning models, deep learning models can effectively extract useful information from huge volume of data to overcome the challenges in signal frame illumination estimation, including various effecting factors, such as camera parameters, scene lighting, object material properties etc. According to (Park and Yoon, 2019), specially, semantic segmentation techniques can present the indoor scene in a more intuitive way for illumination type identification. Identifying the illuminant invariant features can in turn improve the segmentation accuracy especially in low light indoor scenes. From semantic segmentation models, the 3D indoor scene can also be reconstructed from dual camera system by utilizing the illumination reflectance hints for stereo matching. This technique can be used for improving ground vehicles' autonomous navigation performance under the low-light environments, a challenging task for only utilizing traditional machine learning and image processing algorithms.

Utilizing different light sources (e.g. fluorescent lamps, lighting emitting diode (LED) bulbs etc.) with different illumination parameters (e.g. power, frequencies), visible light communication (VLC) has become a strong candidate for indoor communication and indoor positioning as they can provide services wherever lighting is needed, and these systems do not need any extra power for data generation other than the necessary signal processing for illumination estimation. VLC systems will also not generate any radio frequency (RF) interference and therefore could be deployed in environments where RF radiation is strictly restricted. In addition, compared to radio waves rivals, VLC systems provide better positioning accuracy, since they suffer less from multipath effects as well as interference from other wireless handsets (Zhang and Kavehrad, 2013). Machine learning models

can be integrated into the visible light positioning system to improve the performance of indoor positioning in both 2D and 3D spaces, whose results can meet out the standards needed in most designated indoor positioning applications (Tran and Ha, 2022). Outdoor illumination affects the appearances of scene in more complex ways. Weak cues, such as the sky, the vertical surfaces, the ground, and the convex objects in the image, can be extracted from different portions of the image, to estimate the illumination from the single image. Additional information like GPS coordinates, camera orientation and time etc. are tagged in the images, which can further constrain the position of the sun and increase confidences for illumination estimation probability maps (Lalonde et al., 2012).

Even illumination estimation is useful for many downstream applications, removing the illumination effects can also benefit many other computer-vision task. For example, optical flow technique is lately re-gaining the attention in the community for object tracking and speed estimation purposes (Shah and Xuezhi, 2021). Despite the numerous and effective methods published during the last decades, optical flow determination remains challenging due to the data variability arising from both the acquisition conditions and the scenes themselves. (Trinh and Daul, 2019) propose an optical flow approach , that designs a wide variety of new illumination-invariant descriptors via identifying illumination condition changes for weak texture determination and scene motion estimations. Image dehazing is a common post-processing technique to remove the illumination effect for image visual quality improvement, which is widely used in automatic driving and video surveillance. (Chen et al., 2022) proposed a dehazing detection network (DDNet) which receives RGB images and inverse-intensity chromaticity (IIC) images as inputs of a backbone network, EfficientNet-B0, to extract features. The IIC transformation generates a new color space, which can be used to describe the illumination conditions of dehazed images that loses its illumination consistency. An adaptive feature fusion method, is used to fuse the learned RGB and IIC features, thereby improving dehazing region detection capability. The adaptive feature fusion scheme adaptively adjusts the weights in terms of the contribution of each stream.

Table 1 listed published review articles related to the topics of illumination estimation, color constancy and related applications. The article keywords and main contributions have also been summarized in the table. Table 2 listed published research articles related to the above topics, with summarized innovations, outcomes, evaluation metrics. For evaluation metrics, angular error and mean squared error are most commonly used, which represent the cosine angular distance and the Euler distance between the predicted and ground truth color vectors.

Table 1. Published review articles about illumination estimation, color constancy algorithms and related applications.

Title of Study	Keywords	Outcome/Significance
Modeling Scene Illumination Color for Computer Vision and Image Reproduction: A survey of computational approaches (Barnard, 1998)	Illumination modeling, Image reproduction and Image enhancement.	Discussion on the progress in modeling the scene illumination that will result in progress in computer vision, image enhancement, and image reproduction. Also, the nature of image formation and acquisition, overview of the computational approaches have also been investigated.
Computational Color Constancy: Survey and Experiments (Gijsenij et al., 2011)	Statistical and learning based color constancy methods	Various publicly available methods for computational color constancy, of which some are considered to be state-of-the-art, are evaluated on two data sets (the grey-ball SFU (Simon fraser university -set and the Color-checker-set) .
Machine learning in indoor visible light positioning systems: A	Machine learning; illumination-based positioning algorithms	Deep discussions of articles published during the past five years in terms of their proposed algorithm, space (2D/3D), experimental

review (Tran and Ha, 2022)		simulation method), positioning accuracy, type of collected data, type of optical receiver, and number of transmitters.
Deep Neural Models for Illumination Estimation and Relighting: A Survey (Einabadi et al., 2021)	Deep learning; Illumination Estimation and Relighting	Discussion on the main characteristics of the current deep learning methods, dataset and possible future trends for illumination estimation and relighting.
Deep learning for monocular depth estimation: A review (Ming et al., 2021)	Deep learning; Monocular depth estimation	Summarize and categorize the deep learning models for monocular depth estimation. Publicly available datasets and the corresponding evaluation metrics are introduced. The novelties and performance of these methods are compared and discussed.
Single image depth estimation: An overview (Mertan et al., 2022)	Illumination-based depth estimation; Single image depth estimation problem.	Investigations into the mechanisms, principles, and failure cases of contemporary solutions for depth estimation.

Table 2. Published review articles about illumination estimation, color constancy algorithms and related applications.

Title of Study	Methodology/ Technique used	Performance metric	Outcome/Significance
Machine learning approach to color constancy (Agarwal et al., 2007)	Ridge regression and Support Vector Regression	Uncertainty analysis	The shorter training time and single parameter optimization provides a potential scope for real time video tracking application
Convolutional Color Constancy (Barron, 2015)	Discriminative learning using convolutional neural networks and structured prediction	Angular error (mean, median, tri-mean, Error for lowest/ highest 25% predictions	The model can improve performance on standard benchmarks (like White-patch, grey-world etc) by nearly 40%
CNN-Based Illumination Estimation with Semantic Information (Choi et al., 2020)	CNN with new pooling layer to distinguish between useful data and noisy data, and thus efficiently remove noisy data during learning and evaluating.	Mean angular error (MSE) (mean, median, tri-mean, Error for lowest/ highest 25% predictions	Takes computational color constancy to higher accuracy and efficiency by adopting a novel pooling method. Results prove that the proposed network outperforms its conventional counterparts in estimation accuracy
Color Constancy Using CNNs (S. Bianco et al., 2015)	CNN (max pooling, one fully connected layer and three output nodes.)	Angular error (the minimum 10th percentile, 90th percentile, median, maximum, minimum) (Comparison over the methods)	Integrate feature learning and regression into one optimization process, which leads to a more effective model for estimating scene illumination. Improves stability of the local illuminant estimation ability of the proposed method
Deep Learning-Based Computational Color Constancy with	CMoDE fusion technique, multi-stream	Angular error (mean, median, tri-mean,	CMoDE-based DCNN brings significant progress towards efficiency of using computing

Convoluted Mixture of Deep Experts (CMoDE) Fusion Technique (H. -H. Choi and B. -J. Yun, 2020)	deep neural network (MSDNN)	mean of best/worst 25%) (Comparison over the methods)	resources, as well as accuracy of estimating illuminants
Fast Fourier Color Constancy (J. T. Barron and Y. Tsai, 2017)	Fast Fourier Color Constancy in Frequency domain	MSE (mean, median, tri-mean, best/worst 25%)	The method operates in the frequency domain, produces lower error rates than the previous state-of-the-art by $13 - 20\%$ while being $250 - 3000\times$ faster
Color Constancy by Deep Learning (Lou et al., 2015)	DNN-based regression	Angular error (mean, median, standard deviation)	The method outperforms the state-of-the-art by 9%. In cross dataset validation, this approach reduces the median angular error by 35%. The algorithm operates at more than 100 fps during testing.
As-Projective-As-Possible Bias Correction for Illumination Estimation Algorithms (Afifi et al., 2019)	Improving the accuracy of the fast statistical-based algorithms by applying a post-estimate bias-correction function to transform the estimated R, G, B vector such that it lies closer to the correct solution.	MSE (median, tri-mean, best/worst 25%)	Propose an As-projective-as-possible (APAP) projective transform that locally adapts the projective transform to the input R, G, B vector which is effective over the state-of-the-art statistical methods
Robust channel-wise illumination estimation (Laakom et al., 2021)	Efficient CNN	Angular error (mean, median, tri-mean, best/worst 25%)	The method substantially reduces the number of parameters needed to solve the task by up to 90% while achieving competitive experimental results compared to state-of-the-art methods
Deep Outdoor Illumination Estimation (Hold-Geoffroy et al., 2017)	CNN for outdoor illumination estimations	MSE; scale-invariant MSE; per-color scale invariant MSE.	An extensive evaluation on both the panorama dataset and captured HDR environment maps shows significantly superior performances
Fast Spatially-Varying Indoor Lighting Estimation (Garon et al., 2019)	CNN for indoor illumination estimations	MSE Mean Absolute Error	Achieve lower lighting estimation errors and are preferred by users over the state-of-the-art models
Monte Carlo Dropout Ensembles for Robust Illumination Estimation (Laakom et al., 2020)	Monte Carlo dropout	Angular error (mean, median, tri-mean, best/worst 25%)	The proposed framework leads to state-of-the-art performance on INTEL-TAU dataset

Very Deep Learning-Based Illumination Estimation Approach with Cascading Residual Network Architecture (CRNA) (Choi and Yun, 2021)	Cascading Residual Network Architecture (CRNA), which incorporates the ResNet and cascading mechanism into the deep convolutional neural network (DCNN)	Angular error (mean, median, tri-mean, best/worst 25%) (Comparison over the methods)	The proposed approach delivers more stable and robust results and implies the generalization potential for deep learning models across different applications by comparative experiments in different datasets.
Effective Learning-Based Illuminant Estimation Using Simple Features (Cheng et al., 2015)	A learning-based method based on four simple color features and show how to use this with an ensemble of regression trees to estimate the illumination	Angular error (mean, median, tri-mean, best/worst 25%)	A learning-based method based on four simple color features and propose how to use this with an ensemble of regression trees to estimate the illumination (develop a learning-based illumination estimation method with a running-time of statistical methods.)
On deep learning techniques to boost monocular depth estimation for autonomous navigation (de Queiroz Mendes et al., 2021)	A lightweight and fast supervised CNN architecture combined with novel feature extraction models which are designed for real-world autonomous navigation	the Scale invariant Error, Absolute Relative Difference, Squared Relative Difference, Log10, Mean Absolute Error, Linear RMSE and Log RMSE with indoor and outdoor ablation studies	Able to determine optimal training conditions, using different deep learning techniques, as well as optimized network structures that enabled the generation of high-quality predictions in reduced processing time.
Learning HDR illumination from LDR panorama images (Jin et al., 2021)	CNN combining with physical modelling	MSE loss function	Results show that the method can predict accurate spherical harmonic coefficients, and the recovered luminance is realistic.
LISU: Low-light indoor scene understanding with joint learning of reflectance restoration (Zhang et al., 2022)	CNN based novel cascade network to study semantic segmentation in low-light indoor environments	Overall Accuracy, Mean accuracy, Mean intersection over Union (mIoU)	The approach is compared with other CNN-based segmentation frameworks, including the state-of-the-art DeepLab v3+, on the proposed real data set in terms of mIoU, the experimental results also show that the semantic information supports the restoration of a sharper reflectance map, thus further improving the segmentation.
GeoRec: Geometry-enhanced semantic 3D reconstruction of RGB-D indoor scenes (Huan et al., 2022)	Geometry-enhanced multi-task learning network	Mean chamfer distance error, Mean average precision mAP,	With the parsed scene semantics and geometries, the proposed GeoRec reconstructs an indoor scene by placing reconstructed

		Mean intersection over Union mIoU	object mesh models with 3D object detection results in the estimated layout cuboid.
DeepLight: light source estimation for augmented reality using deep learning (Kán and Kafumann, 2019)	Residual Neural Network (ResNet)	Angular error	An end-to-end AR system is presented which estimates a directional light source from a single RGB-D camera and integrates this light estimate into AR rendering.
Deep Spherical Gaussian Illumination Estimation for Indoor Scene (Li et al., 2019)	CNN with extra glossy loss function	Peak Signal to noise Ratio (PSNR) and structural similarity	The proposed approach outperforms the state-of-the-arts both qualitatively and quantitatively.
GMLight: Lighting Estimation via Geometric Distribution Approximation (F. Zhan et al., 2022)	Regression network with spherical convolution with a generative projector for progressive guidance in illumination generation	RMSE and scale invariant RMSE	GMLight achieves accurate illumination estimation and superior fidelity in relighting for 3D object insertion
Deep Graph Learning for Spatially Varying Indoor Lighting Prediction (Bai et al., 2022)	A new lighting model (dubbed DSGLight) based on depth-augmented Spherical Gaussians (SG) and a Graph Convolutional Network (GCN)	PSNR & Qualitative Analysis	DSGLight combines both learning and physical models, and encodes both direct lighting and indirect environmental lighting more faithfully and compactly.
Outdoor illumination estimation via all convolutional neural networks (Zhang et al., 2021)	CNN	Angle error	Pruning and quantization are used to compress the network, resulting in a significant reduction in the number of network parameters and the storage space, only with a slight loss of precision
An illumination estimation algorithm based on outdoor scene classification (Li et al., 2020)	Support Vector Machine (SVM) classifiers with optimization algorithm	Angle error (average, median) and reproduction angle error	Achieve state-of-art results in outdoor environment.

4. Barriers in Illumination Estimation and Recommendations to Overcome These Barriers

Even there are many advanced illumination estimations proposed, there are still some challenges in practice because of complex data collection conditions and data internal variations. In the outdoor environment, the number of outdoor scenes considered for model training is usually less than real-world scenes. Including more number of scenes in the dataset will naturally help for model training to reduce the prediction error, and shallow learning methods usually perform better than deep learning methods for small datasets. For video sequences, the temporal coherence can be utilized to smoothen the inferred illumination from prediction models via temporal filtering techniques (Einabadi et al., 2021). In additional, fusing some extra information in the estimation models can also benefit the performance. Semantic segmentation algorithms in RGB or other color

spaces can be integrated into the illumination model to offer the model high frequency local pattern information. Montel Carlo Dropout and pruning methods can also be used to generalize the CNN model with efficient architecture for different datasets considering different well-designed losses (Laakom, 2007, Ferraz 2018). The CNN model can be also integrated with physical model by simultaneously estimating the scene illumination and spherical reflection model which can effectively estimate the intensities and directions of more than one light sources in a scene (Zhan et al., 2020). For heterogeneous material and texture, traditional wavelet transformation, and radial basis transform can be used to reconstruct the high frequency information because of flickering effects and noises (Park et al 2020).

5. Conclusions

An extensive review about illumination estimation and related deep learning models and applications is presented in the study. In the review, the innovations from different advanced illumination estimation models, especially deep learning models, are summarized. Some challenges in adopting illumination estimation models in practice are also pointed out, and some potential solutions to overcome these barriers are offered. It is concluded that deep learning models can be used to process a significant volume of complex data with accurate illumination predictions and depth estimations compared to traditional model and have enabled for real-time applications such as rendering, 3D modelling, and positioning. Some potential future research directions can focus on parameter optimization and model physical explanation to reduce the computational cost and increase the model generalization capabilities. It is anticipated that deep learning based illumination estimation can broadly benefit computer vision community, and also benefit for researchers with vision intelligent needs in different real-world applications.

Author Contributions: Conceptualization, S.S; investigation, S.S. and J.D.; resources, S.S. and J.D.; writing—original draft preparation, S.S; writing—review and editing, S.S. and D.W.; supervision, D.W.; project administration, D.W.; funding acquisition, D.W. All authors have read and agreed to the published version of the manuscript."

References

Afifi, M., Barron, J. T., LeGendre, C., Tsai, Y. T., & Bleibel, F. (2021). Cross-camera convolutional color constancy. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1981-1990).

Afifi, M., Punnappurath, A., Finlayson, G., Brown, M.S., 2019. As-projective-as-possible bias correction for illumination estimation algorithms. J. Opt. Soc. Am. A 36, 71–78. https://doi.org/10.1364/JOSAA.36.000071

Agarwal, V., Gribok, A.V., Abidi, M.A., 2007. Machine learning approach to color constancy. Neural Netw. 20, 559–563. https://doi.org/10.1016/j.neunet.2007.02.004

Bai, J., Guo, J., Wan, C., Chen, Z., He, Z., Yang, S., Yu, P., Zhang, Y., Guo, Y., 2022. Deep Graph Learning for Spatially-Varying Indoor Lighting Prediction.

Barnard, K., 1998. Modeling Scene Illumination Colour for Computer Vision and Image Reproduction: A survey of computational approaches (Ph D Thesis). Simon Fraser University.

Barron, J.T., 2015. Convolutional Color Constancy.

Bianco, S., Ciocca, G., Cusano, C., Schettini, R., 2009. Improving Color Constancy Using Indoor–Outdoor Image Classification. IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 17, 2381–92. https://doi.org/10.1109/TIP.2008.2006661

Cardei, V.C., Funt, B., n.d. Committee-Based Color Constancy 3.

Celik, T., Tjahjadi, T., 2012. Adaptive colour constancy algorithm using discrete wavelet transform. Comput. Vis. Image Underst. 116, 561–571. https://doi.org/10.1016/j.cviu.2011.12.004

Chen, J., Yang, G., Ding, X., Guo, Z., Wang, S., 2022. Robust detection of dehazed images via dual-stream CNNs with adaptive feature fusion. Comput. Vis. Image Underst. 217, 103357. https://doi.org/10.1016/j.cviu.2022.103357

Cheng, D., Price, B., Cohen, S., Brown, M.S., 2015. Effective learning-based illuminant estimation using simple features, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 1000–1008. https://doi.org/10.1109/CVPR.2015.7298702

Choi, H.-H., Kang, H.-S., Yun, B.-J., 2020. CNN-Based Illumination Estimation with Semantic Information. Appl. Sci. 10. https://doi.org/10.3390/app10144806

Choi, H.-H., Yun, B.-J., 2021. Very Deep Learning-Based Illumination Estimation Approach With Cascading Residual Network Architecture (CRNA). IEEE Access 9, 133552–133560. https://doi.org/10.1109/ACCESS.2021.3115942

de Queiroz Mendes, R., Ribeiro, E.G., dos Santos Rosa, N., Grassi, V., 2021. On deep learning techniques to boost monocular depth estimation for autonomous navigation. Robot. Auton. Syst. 136, 103701. https://doi.org/10.1016/j.robot.2020.103701

Einabadi, F., Guillemaut, J.-Y., Hilton, A., 2021. Deep Neural Models for Illumination Estimation and Relighting: A Survey. Comput. Graph. Forum 40, 315–331. https://doi.org/10.1111/cgf.14283

F. Zhan, Y. Yu, C. Zhang, R. Wu, W. Hu, S. Lu, F. Ma, X. Xie, L. Shao, 2022. GMLight: Lighting Estimation via Geometric Distribution Approximation. IEEE Trans. Image Process. 31, 2268–2278. https://doi.org/10.1109/TIP.2022.3151997

Ferraz, C.T., Borges, T.T.N., Cavichiolli, A., Gonzaga, A., Saito, J.H., n.d. Evaluation of Convolutional Neural Networks for Raw Food Texture Classification under Variations of Lighting Conditions 4.

Forsyth, D.A., 1990. A novel algorithm for color constancy. Int. J. Comput. Vis. 5, 5–35. https://doi.org/10.1007/BF00056770

Foster, D.H., 2011. Color constancy. Vision Res. 51, 674–700. https://doi.org/10.1016/j.visres.2010.09.006

Foster, D.H., Nascimento, S.M.C., Amano, K., Arend, L., Linnell, K.J., Nieves, J.L., Plet, S., Foster, J.S., 2001. Parallel detection of violations of color constancy. Proc. Natl. Acad. Sci. 98, 8151–8156. https://doi.org/10.1073/pnas.141505198

Foster, D.H., Reeves, A., 2022. Colour constancy failures expected in colourful environments. Proc. R. Soc. B Biol. Sci. 289, 20212483. https://doi.org/10.1098/rspb.2021.2483

Gardner, M.-A., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C., Lalonde, J.-F., 2017. Learning to Predict Indoor Illumination from a Single Image. ACM Trans. Graph. 36. https://doi.org/10.1145/3130800.3130891

Garon, M., Sunkavalli, K., Hadap, S., Carr, N., Lalonde, J.-F., 2019. Fast Spatially-Varying Indoor Lighting Estimation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 6901–6910. https://doi.org/10.1109/CVPR.2019.00707

Gijsenij, A., Gevers, T., 2007. Color Constancy using Natural Image Statistics. https://doi.org/10.1109/CVPR.2007.383206

Gijsenij, A., Gevers, T., van de Weijer, J., 2011. Computational Color Constancy: Survey and Experiments. IEEE Trans. Image Process. 20, 2475–2489. https://doi.org/10.1109/TIP.2011.2118224

H. -H. Choi, B. -J. Yun, 2020. Deep Learning-Based Computational Color Constancy With Convoluted Mixture of Deep Experts (CMoDE) Fusion Technique. IEEE Access 8, 188309–188320. https://doi.org/10.1109/ACCESS.2020.3030912

Hold-Geoffroy, Y., Sunkavalli, K., Hadap, S., Gambaretto, E., Lalonde, J.-F., 2017. Deep outdoor illumination estimation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 2373–2382. https://doi.org/10.1109/CVPR.2017.255

Huan, L., Zheng, X., Gong, J., 2022. GeoRec: Geometry-enhanced semantic 3D reconstruction of RGB-D indoor scenes. ISPRS J. Photogramm. Remote Sens. 186, 301–314. https://doi.org/10.1016/j.isprsjprs.2022.02.014

J. Park, H. Park, S. -E. Yoon, W. Woo, 2020. Physically-inspired Deep Light Estimation from a Homogeneous-Material Object for Mixed Reality Lighting. IEEE Trans. Vis. Comput. Graph. 26, 2002–2011. https://doi.org/10.1109/TVCG.2020.2973050

J. T. Barron, Y. Tsai, 2017. Fast Fourier Color Constancy, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6950–6958. https://doi.org/10.1109/CVPR.2017.735

Jin, X., Zhu, X., Li, Xinxin, Zhang, K., Li, Xiaodong, Zhang, X., Zhou, Q., Xie, S., Fang, X., 2021. Learning HDR illumination from LDR panorama images. Comput. Electr. Eng. 91, 107057. https://doi.org/10.1016/j.compeleceng.2021.107057

Kán, P., Kafumann, H., 2019. DeepLight: light source estimation for augmented reality using deep learning. Vis. Comput. 35, 873–883. https://doi.org/10.1007/s00371-019-01666-x

Laakom, F., Raitoharju, J., Iosifidis, A., Nikkanen, J., Gabbouj, M., 2020. Monte Carlo Dropout Ensembles for Robust Illumination Estimation.

Laakom, F., Raitoharju, J., Nikkanen, J., Iosifidis, A., Gabbouj, M., 2021. Robust channel-wise illumination estimation.

Lalonde, J.-F., Efros, A.A., Narasimhan, S.G., 2012. Estimating the Natural Illumination Conditions from a Single Outdoor Image. Int. J. Comput. Vis. 98, 123–145. https://doi.org/10.1007/s11263-011-0501-8

Lou, Z., Gevers, T., Hu, N., & Lucassen, M. P. (2015, September). Color Constancy by Deep Learning. In BMVC (pp. 76-1).

Li, M., Guo, J., Cui, X., Pan, R., Guo, Y., Wang, C., Yu, P., Pan, F., 2019. Deep Spherical Gaussian Illumination Estimation for Indoor Scene. https://doi.org/10.1145/3338533.3366562

Li, N., Li, C., Chen, S., Kan, J., 2020. An Illumination Estimation Algorithm based on Outdoor Scene Classification. Int. J. Circuits Syst. Signal Process. 14. https://doi.org/10.46300/9106.2020.14.54

Lou, Z., Gevers, T., Hu, N., Lucassen, M.P., 2015. Color Constancy by Deep Learning, in: Procedings of the British Machine Vision Conference 2015. Presented at the British Machine Vision Conference 2015, British Machine Vision Association, Swansea, p. 76.1-76.12. https://doi.org/10.5244/C.29.76

Luo, Y., Wang, X., Wang, Q., 2022. Which Features Are More Correlated to Illuminant Estimation: A Composite Substitute. Appl. Sci. 12. https://doi.org/10.3390/app12031175

Mertan, A., Duff, D.J., Unal, G., 2022. Single image depth estimation: An overview. Digit. Signal Process. 123, 103441. https://doi.org/10.1016/j.dsp.2022.103441

Ming, Y., Meng, X., Fan, C., Yu, H., 2021. Deep learning for monocular depth estimation: A review. Neurocomputing 438, 14–33. https://doi.org/10.1016/j.neucom.2020.12.089

P. V. Gehler, C. Rother, A. Blake, T. Minka, T. Sharp, 2008. Bayesian color constancy revisited, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. https://doi.org/10.1109/CVPR.2008.4587765

Park, M.-G., Yoon, K.-J., 2019. As-planar-as-possible depth map estimation. Comput. Vis. Image Underst. 181, 50–59. https://doi.org/10.1016/j.cviu.2019.02.001

S. Bianco, C. Cusano, R. Schettini, 2015. Color constancy using CNNs, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 81–89. https://doi.org/10.1109/CVPRW.2015.7301275

Sarker, I.H., 2021. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput. Sci. 2, 420. https://doi.org/10.1007/s42979-021-00815-1

Schaefer, G., Hordley, S., Finlayson, G., 2005. A combined physical and statistical approach to colour constancy. https://doi.org/10.1109/CVPR.2005.20

Shah, S.T.H., Xuezhi, X., 2021. Traditional and modern strategies for optical flow: an investigation. SN Appl. Sci. 3, 289. https://doi.org/10.1007/s42452-021-04227-x

Tan, R. T., Nishino, K., & Ikeuchi, K. (2004). Color constancy through inverse-intensity chromaticity space. JOSA A, 21(3), 321-334.

Tran, H.Q., Ha, C., 2022. Machine learning in indoor visible light positioning systems: A review. Neurocomputing 491, 117–131. https://doi.org/10.1016/j.neucom.2021.10.123

Trinh, D.-H., Daul, C., 2019. On illumination-invariant variational optical flow for weakly textured scenes. Comput. Vis. Image Underst. 179, 1–18. https://doi.org/10.1016/j.cviu.2018.11.004

Van De Weijer, J., Gevers, T., & Gijsenij, A. (2007). Edge-based color constancy. IEEE Transactions on image processing, 16(9), 2207-2214.

Vazquez-Corral, J., Vanrell, M., Baldrich, R., Tous, F., 2012. Color Constancy by Category Correlation. Image Process. IEEE Trans. On 21, 1997–2007. https://doi.org/10.1109/TIP.2011.2171353

Wang, X., Shen, J., 2019. Machine Learning and its Applications in Visible Light Communication Based Indoor Positioning, in: 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS). Presented at the 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), IEEE, Shenzhen, China, pp. 274–277. https://doi.org/10.1109/HPBDIS.2019.8735490

Zhan, F., Zhang, C., Yu, Y., Chang, Y., Lu, S., Ma, F., Xie, X., 2020. EMLight: Lighting Estimation via Spherical Distribution Approximation.

Zhang, K., Li, Xinxin, Jin, X., Liu, B., Li, Xiaodong, Sun, H., 2021. Outdoor illumination estimation via all convolutional neural networks. Comput. Electr. Eng. 90, 106987. https://doi.org/10.1016/j.compeleceng.2021.106987

Zhang, N., Nex, F., Kerle, N., Vosselman, G., 2022. LISU: Low-light indoor scene understanding with joint learning of reflectance restoration. ISPRS J. Photogramm. Remote Sens. 183, 470–481. https://doi.org/10.1016/j.isprsjprs.2021.11.010

Zhang, W., Kavehrad, M., 2013. Comparison of VLC-based indoor positioning techniques, in: Dingel, B.B., Jain, R., Tsukamoto, K. (Eds.), . Presented at the SPIE OPTO, San Francisco, California, USA, p. 86450M. https://doi.org/10.1117/12.2001569