CRF-net: Single Image Radiometric Calibration using CNNs

Han Li
College of William & Mary

Pieter Peers
College of William & Mary

ABSTRACT

In this paper we present CRF-net, a CNN-based solution for estimating the camera response function from a single photograph. We follow the recent trend of using synthetic training data, and generate a large set of training pairs based on a small set of radiometrically linear images and the DoRF database of camera response functions. The resulting CRF-net estimates the parameters of the EMoR camera response model directly from a single photograph. Experimentally, we show that CRF-net is able to accurately recover the camera response function from a single photograph under a wide range of conditions.

CCS CONCEPTS

• **Computing methodologies → Camera calibration**;

KEYWORDS

Radiometric Calibration, Camera Response Function, Convolutional Neural Network

ACM Reference format:
Han Li and Pieter Peers. 2017. CRF-net: Single Image Radiometric Calibration using CNNs. In *Proceedings of 14th European Conference on Visual Media Production (CVMP 2017), London, United Kingdom, December 11–13, 2017 (CVMP 2017),* 9 pages.
https://doi.org/10.1145/3150165.3150170

1 INTRODUCTION

Radiometric calibration is a necessary pre-processing step for many computer vision and computer graphics algorithms that rely on a linear relation between pixel intensities and scene irradiance. While the raw values returned by image sensors are radiometrically linear to observed irradiance, a non-linear tone-mapping (i.e., camera response function) is applied by the camera to produce a photograph that matches human perception of the scene. Many existing radiometric calibration methods require direct control over the capture process (e.g., adding a calibration target [4] or capturing an exposure stack [5, 24, 26]). Once the camera response function is known, radiometric linearization (i.e., undoing the non-linear transformation of the camera response function) is trivial. However, with the proliferation of computer vision methods that rely on community datasets without knowledge of the camera or control on the capture process, robust direct single image radiometric calibration has become indispensable.

In the past decade several ingenious single image radiometric calibration methods have been proposed that rely on carefully selected cues such as color mixtures at edges [22, 23, 28], the (a)symmetry of noise distributions [25, 30], the reflectance properties of human faces [21], and temporal changes during a single exposure [31]. However, these cues are not universally present in all images, and therefore these methods are not practical for large scale image-databases mined from community repositories. Current practice for such large image sets is to apply a gamma correction during pre-processing instead of a full radiometric calibration. While gamma correction is better than directly using tone-mapped pixel intensities, it is still a poor approximation for most camera response functions [28].

In this paper, we propose a more robust single image radiometric calibration method based on convolutional neural networks (CNN), named CRF-net (or Camera Response Function net). The proposed network takes as input a single well-exposed photograph, and outputs an estimate of the camera response function in the form of an 11-parameter EMoR (Empirical Model of camera Response functions) model [10]. For training CRF-net, we rely the DoRF (Database of Response Functions) database of 201 measured camera response functions [10] to synthesize a large set of tone-mapped images from a much smaller set of radiometrically linear images. Moreover, we introduce a simple oracle for predicting which image windows are likely to produce good results. We experimentally validate the accuracy and robustness of the proposed CRF-net.

2 RELATED WORK

Radiometric calibration aims to recover the camera-dependent response function that relates pixel intensities and their underlying relative scene irradiance. Common representations for camera response functions are gamma curves [24], generalized gamma curves [28], non-parametric (i.e., tabulated) functions [5], polynomials [26], a data-driven model [10], and a log-PCA model [16]. Early work on radiometric calibration assumed full control over the camera, and exploited the relative irradiance relation in a series of differently exposed images of a static scene [5, 24, 26]. Subsequent developments have relaxed the strict-alignment requirement [9, 14, 17], allowed for non-static lighting [13], and estimated camera response functions from community photo-collections [6, 7, 19, 27].

All these prior methods rely on multiple input images to estimate the camera response functions. Lin et al. [22] were the first to demonstrate radiometric calibration from a single image by finding the camera response function that minimizes the deviation from the expected linear mixture of colors at edges, and in follow up work [23] on higher order edge information in grayscale images. Ng et al. [28] observe that identifying good edges is difficult, and instead rely on the more general and easier to identify *locally planar irradiance points*. Similar to edge information, Wilburn et al. [31]

CVMP 2017, December 11–13, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5329-8/17/12…$15.00
https://doi.org/10.1145/3150165.3150170

Output Size	Configuration	Short-cut
114×114	$\begin{bmatrix} 7 \times 7 \times 64 \end{bmatrix}$, stride 2	
57×57	max pool 3×3, stride 2	
57×57	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1 \times 256 \end{bmatrix}$
57×57	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \end{bmatrix} \times 2$	identity
29×29	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1 \times 512 \end{bmatrix}$
29×29	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 1$	identity
23×23	average pool 7×7, stride 1	
11	fully connected	

Table 1: Summary of the DeepDetect's [1] ResNet-18 architecture used for CRF-net.

look at the temporal mixing of two irradiances (e.g., linear motion blur). Instead of edge information, Matsushita et al. [25] exploit the asymmetry in the distribution of camera noise introduced when applying the camera response function. However, extracting noise distributions for the whole pixel range is often difficult. Instead of a noise profile, Takmatsu et al. [30] rely on the variance of image noise estimated from uniform colored regions. Recently, Li et al. [21] proposed to exploit the intrinsic properties of human faces –more precisely, the low-rank nature of skin albedo gradients– to estimate the camera response function from a single photograph containing a human face.

The above single photograph radiometric calibration methods require very specific cues to be present in the image in order to accurately estimate the camera response function. These cues are not universally present in all photographs and/or are difficult to reliably detect, potentially resulting in a suboptimal (or failed) radiometric calibration. The proposed CRF-net does not rely on a single specific information cue, and therefore is more generally applicable.

A related, complementary, topic to radiometric calibration is inverse tone-mapping, where the main goal is to recreate a high dynamic range photograph from low dynamic range input, including oversaturated pixels. Concurrent work in inverse tone-mapping has explored CNNs as a tool for inferring plausible radiance values for oversaturated pixels [8, 33], assuming a known camera response function or a particular type of input (e.g., panoramic images with visible sun). The proposed CRF-net is complementary; it ignores oversaturation while recovering the camera response function for general scenes.

3 CRF-NET

3.1 Radiometric Calibration

The goal of radiometric calibration is to recover the camera response function (f) that translates measured scene irradiance (I) to pixel intensities (M):

$$M = f(I). \tag{1}$$

Given the inverse function $g = f^{-1}$, the original radiometrically linear image I can then be recovered (up to a scale factor and ignoring over and underexposed pixels) from the captured photograph.

Estimation of the camera response function from a single photograph is an ill-conditioned problem. We therefore simplify the problem by assuming that the same camera response function is applied to each color channel separately, and that it is the only source of non-linear transformation between the measured scene irradiance and the reported pixel intensities. While, this assumption fails to account for any non-linear effects due to gamut mapping of oversaturated pixels [2, 3, 15, 32], we found it to work well in practice for photographs with a moderate dynamic range (and thus a limited amount of under- and oversaturation). We target radiometric calibration of photographs typically encountered in online photo-collections. The vast majority of these photographs are captured with consumer or cell-phone cameras with auto-exposure and auto-white balancing enabled. We will therefore focus on well-exposed and correctly balanced photographs, as these provide the most useful information for typical computer vision and computer graphics applications that rely on radiometric calibration.

In contrast to prior work that relies on very specific cues to extract the camera response function from a single photograph, we take a learning based approach to discover descriptive features that can predict the camera response function (output) from a single photograph (input). In particular, we leverage convolutional neural networks to perform radiometric calibration. The resulting network, named CRF-net, directly estimates the parameters (i.e., PCA weights) of the EMoR camera response function model [10]. A CNN based solution could also directly output a radiometrically linear image instead of estimating the camera response function and computing the radiometrically linear image in post-processing as we propose. However, while directly generating the radiometric linear image could potentially model other types of (spatially varying) non-linearities, it would consequently also be more susceptible to produce visually distracting spatial artifacts. While more restrictive, estimating a camera response function guarantees a plausible and consistent radiometrically linear image.

3.2 CRF-net Architecture

CRF-net follows the powerful ResNet-18 architecture [11] from the DeepDetect library [1] implemented in Caffe [12]. This architecture differs from the 18-layer ResNet introduced by He et al. [11]; DeepDetect's ResNet-18 architecture is a cut-out of He et al.'s 50-layer ResNet-50. We opt for this architecture because, based on prior work in single image radiometric calibration, we expect mostly local pixel relations (e.g., edge information) to inform radiometric calibration. Furthermore, a shallow network with small filters also reduces the amount of required training data, greatly facilitating training. Table 1 summarizes the architecture. We also experimented with other network architectures such as VGGnet [29] and AlexNet [18], but these architectures did not produce good results. We add a fully connected layer on top of ResNet-18 that outputs the weights of the 11 largest PCA components of the EMoR model. While Grossberg and Nayar [10] report that a 3 parameter EMoR model already covers 99.5% of the energy, we opt to use an 11 parameter model

as this produces nearly perfect matches on the most challenging camera response functions in the DoRF database (see [10], Fig. 7).

CNNs are often restricted to input images of limited resolution. Likewise, the proposed CRF-net also only operates on 227×227 pixel windows. However, radiometric calibration typically deals with much larger images. We therefore select and separately process 10 well-chosen 227×227 windows from the input image, and aggregate the corresponding estimates. Note that we cannot simply scale the input images to a smaller resolution because the non-linearity of the camera response function would destroy the relation between the (averaged) pixel intensities and the corresponding (averaged) scene irradiance: $I_1 + I_2 = g(M_1) + g(M_2) \neq g(M_1 + M_2)$. However, an ill-chosen 227×227 window (e.g., covering only the sky) will also not produce good results. We posit that a "good" window should cover a large range of (union of) red, green, and blue intensities. We therefore repeatedly select and test random candidate windows, until we have found 10 windows whose pixel value histograms (i.e., a single histogram per window aggregating all color channel values in 256 bins) contain at least 220 non-empty bins each. If after a certain number of attempts no such windows are found, then we select the 10 windows that best covered the intensity range. However, in such a case we expect a suboptimal radiometric calibration. Finally, we aggregate the estimated camera response functions from the 10 well-chosen windows, by removing the outliers and averaging the corresponding parameters of the remaining estimated camera response functions.

3.3 Training

Radiometric calibration is significantly different from other problems, such as object recognition, intrinsic decomposition, etc., on which CNNs have successfully been applied. Therefore, we cannot refine an existing network. Consequently, we are forced to train CRF-net from scratch, and thus we require an extensive training dataset. Obtaining a large dataset of photographs for a large variety of scenes and capture conditions, together with corresponding ground truth camera response functions from a diverse set of camera models, is time-consuming and difficult. Instead we follow the recent trend of using synthetic training data.

We have collected a set of 595 well-exposed radiometrically linear "RAW" photographs, captured with 3 different camera models (i.e., Canon EOS 600D, Nikon D800, and Nikon D300S), from a variety of scenes (approximately 60% indoor scenes and 40% outdoor scenes) captured under a variety of conditions (e.g., clear sky, overcast sky, night time, etc.). From this set of radiometrically linear images, we generate corresponding tone-mapped photographs for each of the 201 camera models in the DoRF database [10]. To reduce storage requirements and minimize disk overhead during training, we scale the radiometrically linear image first by an *integer* factor such that the smallest dimension is just larger than 227. We deliberately only apply an integer scale factor such that each (raw) image pixel is only assigned to a single tone-mapped image pixel. Furthermore, since CRF-net requires 227×227 pixel windows (and we cannot scale tone-mapped images), we select 10 well-chosen pixel windows using the same intensity criterion as detailed in subsection 3.2. Furthermore, we desire to train CRF-net for reasonably exposed images, such as those produced by using the auto-exposure function on a consumer camera; severely underexposed or overexposed image are unlikely to contain sufficient information to retrieve the camera response function and/or to extract any meaningful image information. We therefore precompute for each camera response curve in the DoRF database a scale factor 's_{crf}' that generally produces well-exposed images, roughly approximating the effect of 'auto-exposure'. To further avoid biasing CRF-net to relate overall brightness and the camera response function, we produce 5 slightly different exposed versions by randomly sampling an effective exposure in the range: $[s_{crf} - 0.4, s_{crf} + 0.4]$. In total, our training dataset consists of $595 \times 201 \times 10 \times 5 = 5,979,750$ image windows with corresponding camera response functions.

In addition to the training dataset, we also generate a validation dataset, but using a different set of 20 radiometrically linear RAW images captured with 3 different camera models, of which one is shared with the training dataset (i.e., Canon EOS 600D), and two are new camera models (i.e., Nikon D700 and Canon EOS 60D). We generate synthetic photographs from this set of 20 images using again all 201 camera models from the DoRF database and with 5 different exposures selected in a similar fashion as for the training dataset.

As noted before, we train CRF-net from scratch. However, we found that directly training CRF-net fails to converge to a suitable minimum. To provide a better starting point for training, we first train a slightly different variant that instead of outputting the EMoR parameters, outputs a likelihood that a photograph was generated by each of the camera response functions in the DoRF database (i.e., a classification network where the fully connected layer outputs 201 likelihoods instead of the weights of the 11 PCA components). Due to the similarity of many camera response functions in the DoRF database, the accuracy of this classification network is poor (only 26% of the photographs are correctly classified). However, it serves as a better starting point to refine the full CRF-net. We train the classification network using the standard classification loss function, and with the following hyperparameters: learning policy "step", base learning rate of 0.01, a step size of $500,000$, $2,000,000$ maximum number of iterations, momentum 0.9, and a weight decay of 0.0005. After convergence, we replace the fully connected layer of the classification network, copy the trained CNN parameters, and directly use the Euclidian distance between the estimated and ground truth PCA weights of the camera response function as the loss function. We use the same training images and hyperparameters to refine CRF-net from the classification network, except for: base learning rate 0.0001, step size $20,000$, and maximum number of iterations $25,000$.

4 RESULTS

We will employ two kinds of error metrics to gauge the accuracy of the recovered camera response functions. The "estimation error" is defined as the L2 distance between two normalized curves (discretized in 1024 relative radiance samples as in the EMoR PCA basis). As we represent the camera response functions using the EMoR PCA model, we can simply, due to the orthogonality of the PCA basis, use the equivalent L2 distance between the corresponding PCA weights. This error relates to RMSE of the camera response functions as: $L2 = \sqrt{1024} \times RMSE$. While the estimation

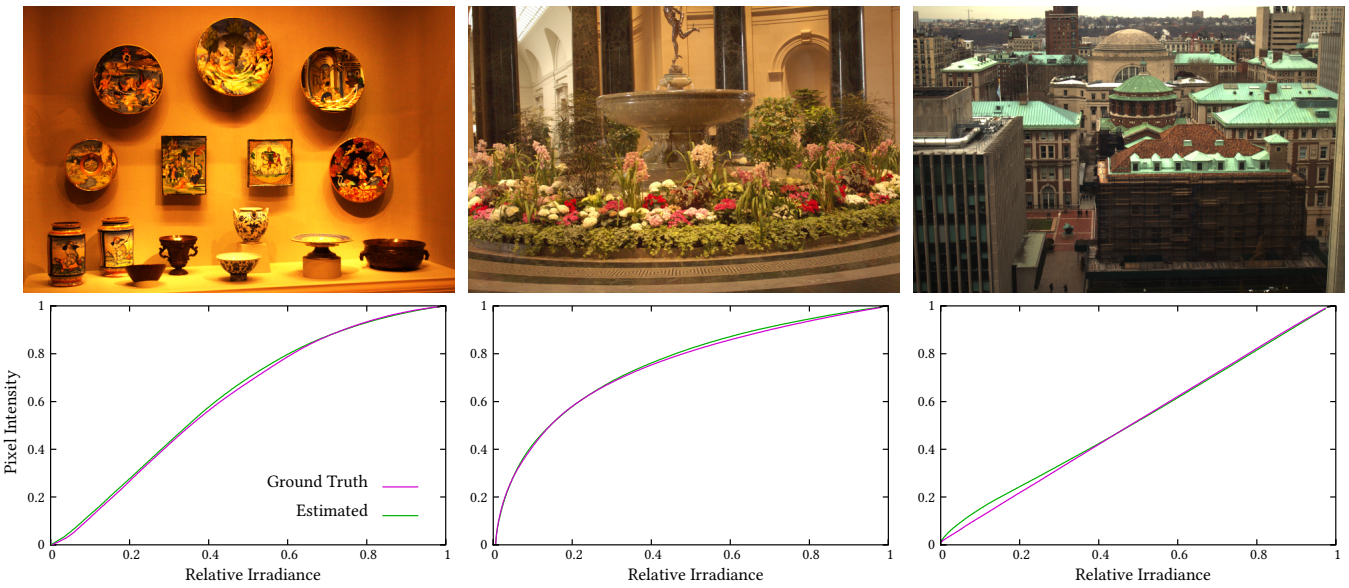

Figure 1: Three examples of photographs and corresponding ground truth (purple) and estimated (green) camera response functions. The estimation errors are: 0.326, 0.267, 0.491, **and the linearization errors are (**$\times 10^{-2}$**):** 0.649, 0.835, 1.482.

error indicates how similar both camera response curves are, it does not take in account whether the whole range is meaningful with respect to the target image. For example, the error outside the range of pixel values present in the image has little influence on the accuracy at which the image can be radiometrically linearized. We therefore also consider the "linearization error" that is defined as the RMSE between the images linearized by the ground truth and estimated camera response functions. We ensure that the peak signal in the ground truth image equals 1 (and scale the linearized image accordingly). Hence, the reported RMSE relates to PSNR as: $-20\log_{10}(RMSE)$.

Figure 1 shows three images generated by applying a camera response function from the DoRF database to a radiometrically calibrated image not part of the training dataset. In addition we show the ground truth (purple) and recovered camera response functions (green) which are a close match. When the image contains many oversaturated pixels or a large contrast, it becomes difficult to find many good windows (Figure 2), resulting in a less accurate radiometric calibration. Depending on the application, the resulting camera response functions and/or radiometrically linearized images might still be of sufficient quality. Over the full validation dataset, the average estimation and linearization errors are 1.607 and 2.089$\times 10^{-2}$ respectively.

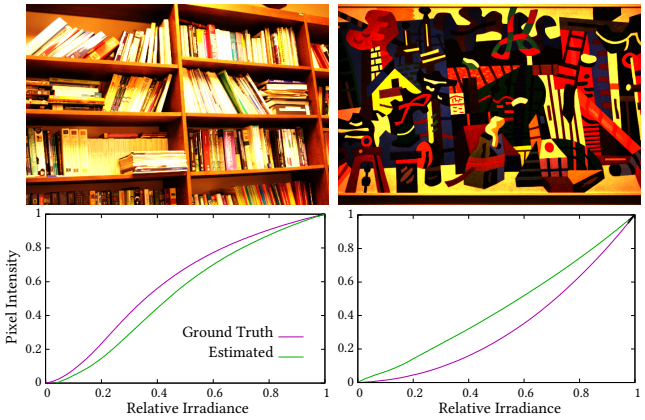

Figure 2: Examples of suboptimal radiometric calibration. The left image exhibits many oversaturated pixels, whereas the right exhibits a very high contrast. In both cases, it is difficult to find good windows that sufficiently (and uniformly) cover the full pixel range. The respective estimation (and linearization ($\times 10^{-2}$)**) errors are:** 2.365 (3.037) **and** 3.925 (7.485)**.**

5 DISCUSSION

CRF-net only operates on a small 227×227 window, and the content of a window greatly affects the quality of the radiometric calibration. While we aggregate the estimates from 10 windows, it is still instructive to know what kind of windows provide good estimates, and how effective our selection criterion works in practice. Figure 3 compares the camera response curves estimated from a randomly selected window (marked in red) and a window that matches our selection criterion (marked in green). As expected, the random window that exhibits little pixel variations does not provide sufficient cues to estimate an accurate camera response function.

To better understand the limitations of CRF-net, we furthermore validate its robustness against the following factors: variations in exposure, image/feature scale, color vs. grayscale, measurement noise, sharpness/blur, and real photographs (i.e., non-synthetic images).

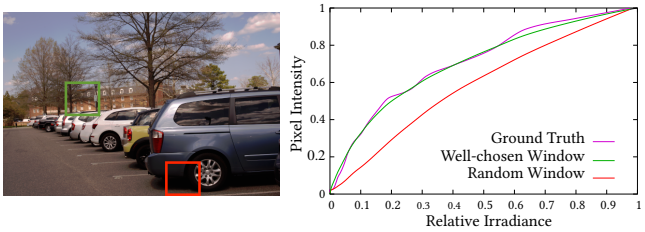

Figure 3: Estimated camera response curves from a single window: a randomly selected one (red) and one selected with the proposed selection criterion (green).

Exposure	0.6	0.8	1.0	1.2	1.4
Estimation	1.621	1.611	1.607	1.647	1.676
Linear. ($\times 10^{-2}$)	2.551	2.331	2.089	2.792	2.999

Table 2: Estimation and linearization errors over the validation dataset for different exposures scaled relatively with respect to the 'ideal' auto-exposure.

5.1 Exposure

We scale the radiometrically linear input images of our validation set by $[0.6, 0.8, 1.0, 1.2, 1.4]$ and compute the evolution of the estimation and linearization errors for different exposure scales (Table 2). From this we conclude that CRF-net is robust for moderate deviations from the optimal exposure, as long as there are windows that cover a sufficiently large range of pixel intensities uniformly. Unless severe, oversaturation only affects local regions and thus we can still find good windows for recovering the camera response function. Undersaturation, on the other hand, typically affects the overall brightness of the whole image, making it difficult to find good windows. Consequently, CRF-net is more sensitive to undersaturation. Camera response functions that tend to overly boost the contrast of the image (e.g., Figure 2, right) suffer from a similar problem as undersaturation. Unlike undersaturation, there exist windows that fulfill our selection criteria. However, the histograms for these windows exhibit a severely skewed distribution, and thus provide insufficient information for certain regions of the intensity range to reliably estimate the camera response function.

5.2 Scale

To ensure CRF-net is not overtrained for a specific image-feature size, we compute the estimation and linearization error on the validation dataset, with double and quadruple resolution. As the images in the validation and training dataset are synthesized by downscaling the original captured RAW images by at least a factor 8, we can easily generate artifact-free higher resolution versions of the corresponding validation images directly from the original captured images (instead of upsampling the images from the validation set). The average estimation (and linearization) errors (1.864 (2.701×10^{-2}) and 2.144 (3.029×10^{-2}) for 2× and 4× respectively) compared to the unscaled errors (1.607 (2.089×10^{-2})) increase slightly. We suspect that the slight increase in error is due to a smaller coverage of the relative scene area with increasing resolution (the window becomes relatively smaller compared to the depicted scene), and thus the variety in observed pixel intensities decreases.

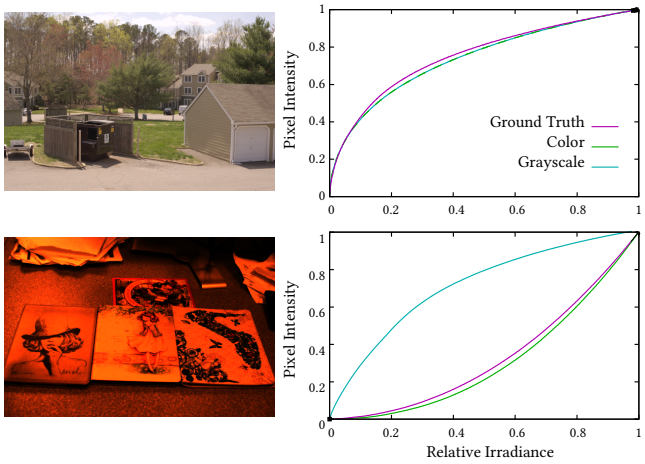

Figure 4: Radiometric calibration of colored versus grayscale images. Grayscale images exhibit less variation in intensity distributions, and are therefore less robust to calibrate. The top row shows a successful calibration for both color and grayscale; the bottom row shows an example where radiometric calibration on a colored image succeeds, but fails on the same grayscale image.

5.3 Grayscale

Inspired by Lin et al. [23], we also validate whether CRF-net requires colored input. By removing the color information, we also remove a significant amount of information for CRF-net to exploit. Furthermore, the training dataset does not contain monochromatic images, and hence CRF-net needs to extrapolate from the learned model to process grayscale images. Figure 4 shows a comparison between two different response functions applied to the color and grayscale version of two images. In both cases, a successful calibration is achieved for the color images. However, the calibration on the grayscale versions of the same images with the same camera response function is bimodal: succeeding in one case without loss of accuracy, and failing on the second case. The average estimation (linearization) error on the validation set are 1.607 (2.089×10^{-2}) for the color input, and 3.546 (4.028×10^{-2}) for the corresponding grayscale versions.

5.4 Measurement Noise

Inspired by prior work that exploits the symmetry of noise distributions [25, 30], we also validate the robustness of CRF-net with respect to noise. For each image in the validation dataset, we add normal distributed noise before applying the camera response curve. Table 3 shows the respective errors for increasing noise variances. These results show a degradation of the calibration accuracy with increasing magnitude of camera noise. In general we observe that when the noiseless calibration is very accurate, camera noise impacts the radiometric calibration to a lesser degree than for cases where the noiseless calibration is less accurate.

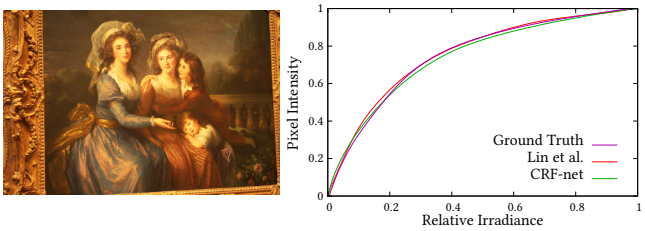

Figure 5: Comparison to between Lin et al.'s single image radiometric calibration method [22] and CRF-net on an image for which the former works well.

Noise σ^2	0	0.5	1	2	4
Estimation	1.607	2.302	3.036	4.726	6.069
Linear. ($\times 10^{-2}$)	2.089	3.359	4.419	9.128	48.20

Table 3: Estimation and linearization error over the validation dataset for different amounts of normal distributed camera noise.

Blur σ^2	0	1	2	4	8
Estimation	1.607	1.728	2.052	2.219	2.335
Linear. ($\times 10^{-2}$)	2.089	2.821	3.398	3.594	3.544

Table 4: Estimation and linearization error over the validation dataset for different amounts of blur.

5.5 Image blur

Depending of the aperture setting, or motion in the scene, certain parts of the image might be blurred. To validate the robustness against blur, we apply differently sized blur filters to the radiometrically linear validation images before applying the camera response function (Table 4). From this experiment we can conclude that our method is not sensitive to moderate amounts of blur, and robust to strong blurring. This seems to suggest that CRF-net only weakly relies on edge information (in contrast to [20, 22, 23]).

5.6 Comparison Prior Work

Our experiments show that CRF-net can robustly estimate the camera response function under a wide range of conditions. A fair comparison to prior work is difficult as it is easy to find examples on which prior single image radiometric calibration methods fail. Nevertheless, even a partial comparison is still instructive to better understand the advantages and limitation of CRF-net. Figure 5 compares the estimated camera response function using CRF-net with that obtained using the method of Lin et al. [22] on a carefully selected photograph for which the latter works well; we found that Lin et al.'s method did not perform well for many examples in our validation set. This example demonstrates that under conditions favorable to Lin et al.'s method, the proposed CRF-net produces comparable or better results.

Currently, for linearizing large image datasets, a simple but robust gamma correction is often favored instead of existing advanced single image radiometric calibration methods. To compare the accuracy of CRF-net to gamma correction, we compute the estimation (and linearization) error using both methods on the validation dataset. The average errors are 1.607 (2.089×10^{-2}) for CRF-net versus 3.132 (5.821×10^{-2}) for gamma correction. The estimation error of CRF-net was lower in 78% (or 86% for the linearization error) of the examples in the validation set compared to naive gamma correction. This clearly demonstrates that CRF-net is a robust and more accurate alternative to gamma correction.

5.7 Non-synthetic Photographs

All our training data and validation data are synthetically generated from radiometrically linear photographs captured using 3 different cameras. This raises the question whether CRF-net is overtrained to the characteristics (e.g., noise) of the image sensors in these cameras. Furthermore, all our synthetically generated images lack the typical post-processing steps camera manufacturers apply to make the photograph "look good". This also raises the question whether CRF-net is robust to such post-processing steps. We validate its robustness to these issues by demonstrating the recovery of the camera response functions from 8 well-exposed captured photographs. The "ground truth" camera response functions are computed from a 3-photograph exposure stack (1 F-stop separation) for each scene using a commercial implementation of the method of Debevec and Malik [5]. To offer a fair comparison (i.e., same dynamic range, same white-balancing, etc.), we compute the corresponding "ground truth" linear image by applying the ground truth camera response function to the selected well-exposed captured photograph. Since we directly use the camera-produced tone-mapped photographs as an input, unknown post-processing is included. Furthermore, none of the camera models are present in the DoRF database. As demonstrated in Figure 6, CRF-net exhibits a similar performance on post-processed non-synthetic photographs as on the synthesized images in the training and validation datasets

While CRF-net is primarily aimed at photographs captured with auto-settings (e.g., exposure, aperture, white-balance, etc.), it is nevertheless instructive to analyze the robustness of CRF-net with respect to variations in these camera parameters. For simplicity, we perform this analysis on a fixed scene (Figure 7 and Figure 8) captured with a single camera (i.e., Nikon 700D). In particular, we validate the robustness with respect to aperture (3.2 to 16.0) , ISO (100 to 6000), lenses (50mm vs. 105mm), exposure (1/200 to 0.8 seconds) and white balance. Variations in aperture, ISO, and lenses did not affect the estimation of the camera response functions.

As expected, CRF-net failed to produce good estimates from severely underexposed and overexposed photographs. This confirms our earlier analysis on the impact of exposure on synthetic data (subsection 5.1). However, we also noted slight variations in accuracy for well-exposed images (Figure 7). There was no specific pattern or correlation between the exposure and the error on the corresponding estimated camera response function (besides over- and underexposure). We attribute these slight differences between the distribution of brightness values in the 227×227 windows; a slight change in exposure can potentially result in a better distribution. Nevertheless, CRF-net produced acceptable estimates for a significant exposure range (1/8 to 0.8 seconds $\approx$ 3 F-stops).

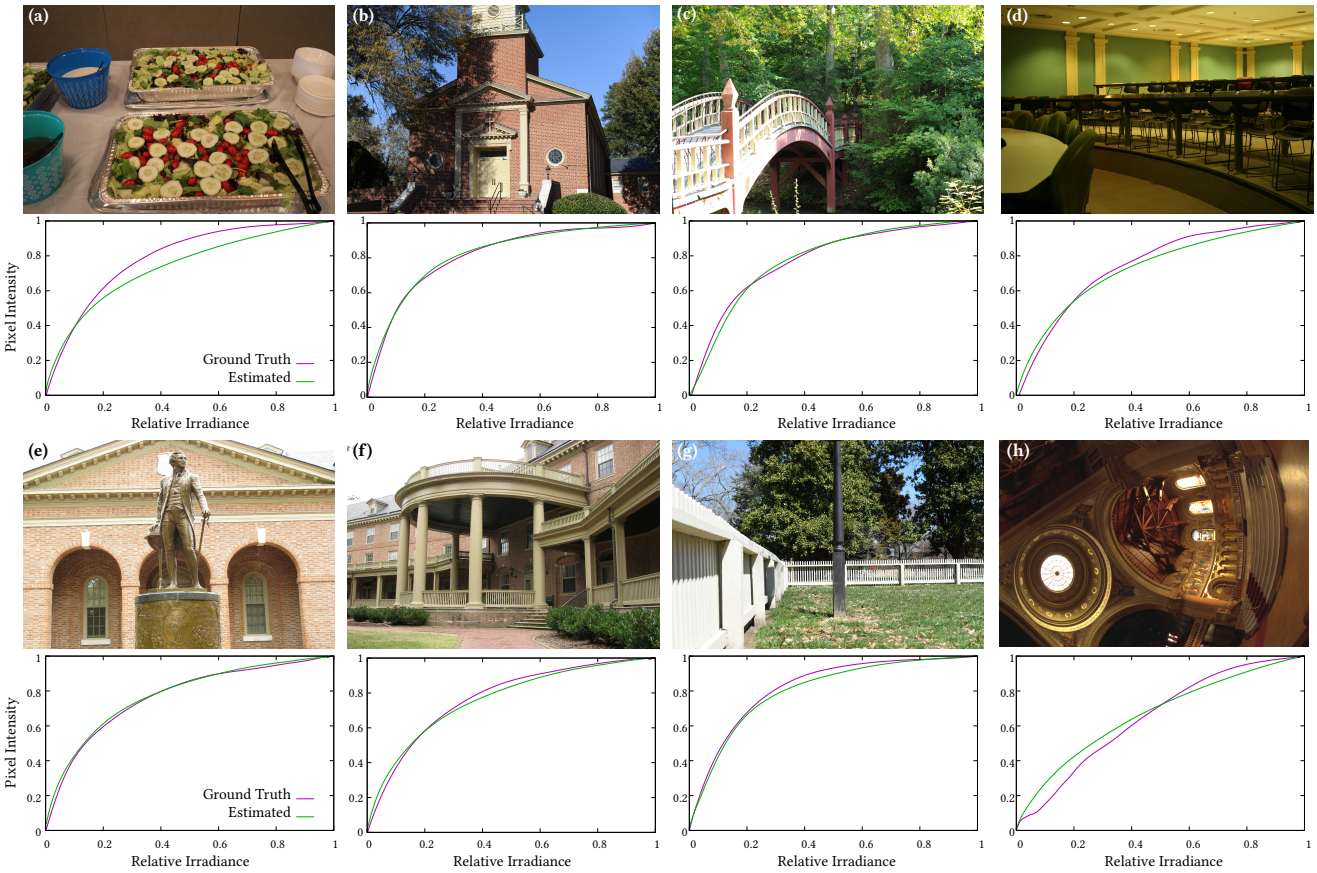

Figure 6: Results from CRF-net applied to a single well-exposed photograph from an exposure stack captured with different camera models (Canon 60D (a), Nikon D700 (b,c), Canon 350D (d), Canon Powershot SX110 (e,f), Nikon D750 (g), and scanned color print film (Fuji 100 ASA, scanned with a Kodak PhotoCD film scanner) obtained from prior work [5] (h). The respective estimation (and linearization ($\times 10^{-2}$)) errors are computed with respect to the response function estimated from the full exposure stack: (a) 2.1758 **(**2.753**), (b)** 0.4221 **(**2.300**), (c)** 0.7224 **(**1.558**), (d)** 1.0366 **(**1.623**), (e)** 0.8234 **(**2.999**), (f)** 0.7927 **(**2.044**), (g)** 0.7517 **(**2.579**), and (h)** 1.7897 **(**6.730**).**

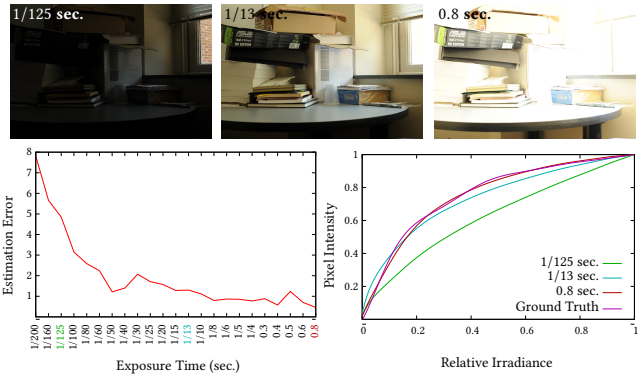

Figure 7: Impact of camera exposure on estimation error. Top: three selected exposures. Bottom Left: estimation error. Bottom Right: estimated camera response functions of the three images shown at the top.

We found that white balance settings significantly impact the ability of CRF-net to estimate the camera response function. CRF-net failed to produce a meaningful estimate for white balance settings that produced an overall cooler tone (Figure 8: *"Incandescent"* and *"Sodium Lamp"*). However, neutral and warmer toned photographs (Figure 8: *"Flash"* and *"Cloudy"*), which typically also encompasses auto white balance settings, did not aversely impact the accuracy of the estimates. As with all methods based on CNNs, the accuracy of the result is related to how well the training data spans the target space. When CRF-net is applied to an image-type not represented by the training data (e.g., with different overall color tone), an incorrect camera response function estimate is produced. Grayscale images are another example of applying CRF-net outside the learned space.

5.8 Limitations

CRF-net makes the a-priori assumption that the camera response curve is the only source of non-linearity. This is not the case when non-linear gamut mapping of overexposed pixels occurs; a problem

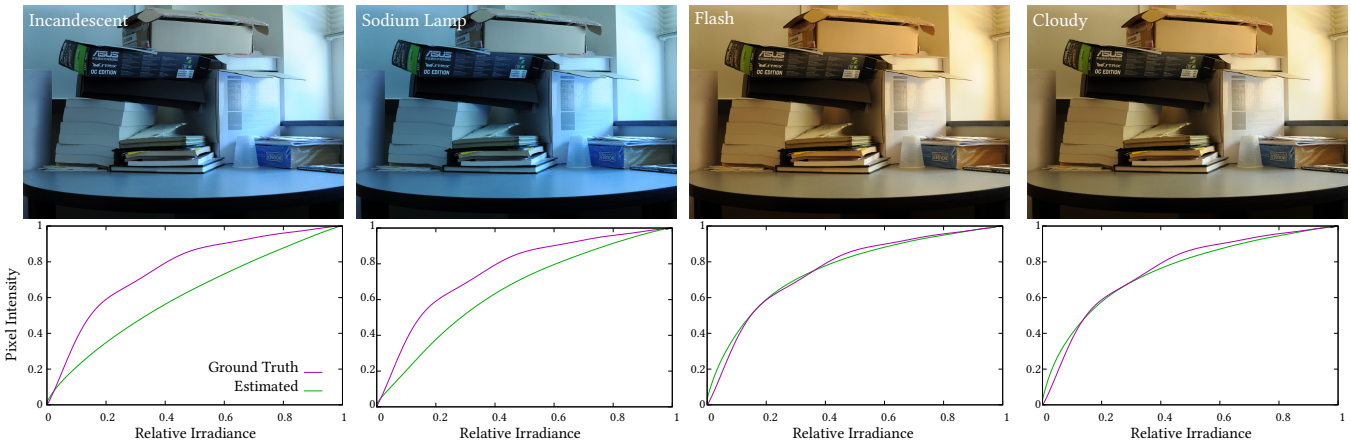

Figure 8: CRF-net fails to recover a meaningful camera response function for white balance settings that produce an overall cooler (i.e., bluish) tone such as the *"Incandescent"* and *"Sodium Lamp"* white-balance settings on this test scene. In contrast, CRF-net produces accurate estimates for white balance settings that produce an overall warmer tone (i.e., reddish) such as the *"Flash"* and *"Cloudy"* white-balance settings.

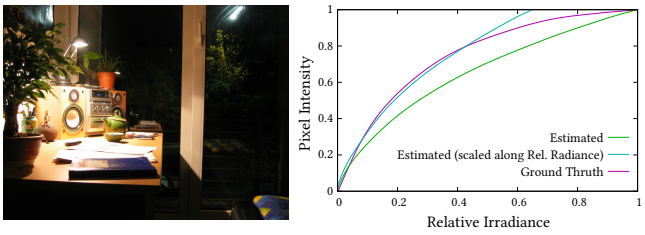

Figure 9: Non-linear gamut mapping of overexposed pixels aversely affects the accuracy of CRF-net. However, ignoring the upper 20% of the pixel range, and scaling the estimated camera response function appropriately, still produces a reasonable approximation.

more common in photographs of scenes with a large dynamic range [2, 3, 15, 32].

Figure 9 shows an example of a scene with a large dynamic range[1]. The estimated camera response curve (green) differs significantly from the ground truth (purple) camera response curve recovered from the exposure stack. However, we observe that estimation of the ground truth camera response function from the exposure stack is unstable, and different results are obtained based on different subsets of the stack. The instabilities seem to be most prominent in the upper 10 to 20 percent of the pixel intensity range. We posit that this is mainly due to non-linearities introduced by gamut mapping of oversaturated pixels. Since the x-axis represents relative irradiance, we can rescale the camera response function along this axis to obtained a better match over the lower 80% of the pixel range (cyan curve). This example suggests that, despite the large dynamic range and non-linear gamut mapping, CRF-net can potentially recover an accurate partial camera response function for a large portion of the range. However, a more extensive analysis is needed to confirm this thesis.

[1]Source: http://resources.mpi-inf.mpg.de/hdr/calibration/pfs.html

6 CONCLUSION

In this paper we presented a CNN-based solution for radiometric calibration from a single input photograph. We have experimentally verified the robustness of CRF-net for a wide range of conditions. We believe CRF-net can serve as a valuable pre-processing step for computer vision and computer graphics algorithms that require a linear relation between pixel intensities and scene radiance on large datasets mined from uncalibrated repositories. For future work, we would like to generalize CRF-net to robustly handle variations in white balance, as well, as grayscale exemplars, by either training CRF-net on single channel images or by extending the training set with different white-balanced versions of the training images similar to how we currently include variations in exposure.

ACKNOWLEDGEMENTS

We wish to thank the anonymous reviewers for their constructive feedback. This work was partially funded by NSF grant IIS-1350323, and a gift from Google.

REFERENCES

[1] Emmanuel Benazera. 2015. DeepDetect. (2015). http://www.deepdetect.com.
[2] Ayan Chakrabarti, Daniel Scharstein, and Todd Zickler. 2009. An Empirical Camera Model for Internet Color Vision. In *BMVC*.
[3] Ayan Chakrabarti, Ying Xiong, Baochen Sun, Trevor Darrell, Daniel Scharstein, Todd E. Zickler, and Kate Saenko. 2014. Modeling Radiometric Uncertainty for Vision with Tone-Mapped Color Images. *PAMI* 36, 11 (2014), 2185–2198.
[4] Young-Chang Chang and J. F. Reid. 1996. RGB calibration for color image analysis in machine vision. *IEEE TIP* 5, 10 (Oct 1996), 1414–1422.
[5] Paul E. Debevec and Jitendra Malik. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *SIGGRAPH '97*. 369–378.
[6] Mauricio Diaz and Peter Sturm. 2011. Exploiting Image Collections for Recovering Photometric Properties. In *CAIP*. 253–260.
[7] Mauricio Díaz and Peter Sturm. 2011. Radiometric Calibration using Photo Collections. In *ICCP*. 1–8.

[8] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał Mantuik, and Jonas Unger. 2017. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* 36, 6 (2017).

[9] M.D. Grossberg and S.K. Nayar. 2003. Determining the Camera Response from Images: What is Knowable? *PAMI* 25, 11 (Nov 2003), 1455–1467.

[10] M.D. Grossberg and S.K. Nayar. 2004. Modeling the Space of Camera Response Functions. *PAMI* 26, 10 (Oct 2004), 1272–1282.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *CVPR*. 770–778.

[12] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In *ICM*. 675–678.

[13] S. J. Kim, J. M. Frahm, and M. Pollefeys. 2008. Radiometric Calibration with Illumination Change for Outdoor Scene Analysis. In *CVPR*.

[14] Seon Joo Kim, David Gallup, Jan-Michael Frahm, and Marc Pollefeys. 2010. Joint radiometric calibration and feature tracking system with an application to stereo. *Computer Vision and Image Understanding* 114, 5 (May 2010), 574–582.

[15] S. J. Kim, H. T. Lin, Z. Lu, S. SÃijsstrunk, S. Lin, and M. S. Brown. 2012. A New In-Camera Imaging Model for Color Computer Vision and Its Application. *PAMI* 34, 12 (Dec 2012), 2289–2302.

[16] Seon Joo Kim and Marc Pollefeys. 2004. Radiometric Alignment of Image Sequences.. In *CVPR*. 645–651.

[17] Seon Joo Kim and Marc Pollefeys. 2008. Robust Radiometric Calibration and Vignetting Correction. *PAMI* 30, 4 (April 2008), 562–576.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *NIPS*. 1097–1105.

[19] S. Kuthirummal, A. Agarwala, D. B Goldman, and S. K. Nayar. 2008. Priors for Large Photo Collections and What They Reveal about Cameras. In *ECCV*. 74–87.

[20] Joon-Young Lee, Yasuyuki Matsushita, Boxin Shi, In-So Kweon, and Katsushi Ikeuchi. 2013. Radiometric Calibration by Rank Minimization. *PAMI* 35, 1 (2013), 144–156.

[21] Chen Li, Stephen Lin, Kun Zhou, and Katsushi Ikeuchi. 2017. Radiometric Calibration From Faces in Images. In *CVPR*.

[22] Stephen Lin, Jinwei Gu, Shuntaro Yamazaki, and Heung-Yeung Shum. 2004. Radiometric Calibration from a Single Image.. In *CVPR*. 938–945.

[23] Steve Lin and L. Zhang. 2005. Determining the radiometric response function from a single grayscale image. In *CVPR*. 66âĂŞ73.

[24] S. Mann and R. W. Picard. 1995. On Being 'undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures. In *Proceedings of IS&T*. 442–448.

[25] Yasuyuki Matsushita and Steve Lin. 2007. Radiometric calibration from noise distributions. In *CVPR*.

[26] T. Mitsunaga and S.K. Nayar. 1999. Radiometric Self Calibration. In *CVPR*, Vol. 1. 374–380.

[27] Zhipeng Mo, Boxin Shi, Sai-Kit Yeung, and Yasuyuki Matsushita. 2017. Radiometric Calibration for Internet Photo Collections. In *CVPR*.

[28] Tian-Tsong Ng, Shih-Fu Chang, and Mao-Pei Tsui. 2007. Using Geometry Invariants for Camera Response Function Estimation. In *CVPR*.

[29] K. Simonyan and A. Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* abs/1409.1556 (2014).

[30] Jun Takamatsu, Yasuyuki Matsushita, and Katsushi Ikeuchi. 2008. Estimating Radiometric Response Functions from Image Noise Variance. In *ECCV*. 623–637.

[31] Bennett Wilburn, Hui Xu, and Yasuyuki Matsushita. 2008. Radiometric calibration using temporal irradiance mixtures. In *CVPR*.

[32] Y. Xiong, K. Saenko, T. Darrell, and T. Zickler. 2012. From pixels to physics: Probabilistic color de-rendering. In *CVPR*. 358–365.

[33] Jinsong Zhang and Jean-François Lalonde. 2017. Learning High Dynamic Range from Outdoor Panoramas. In *ICCV*.