DEEP-LEARNING-BASED INVERSE TONE MAPPING OPERATIONS FOR HDR IMAGE RECONSTRUCTION IN BROADCAST

M. Benyamin

RheinMain University of Applied Sciences, Germany

ABSTRACT

HDR is considered a major topic and will soon replace SDR as production standard. However, most legacy content is only available in lower SDR quality and must always be included into future productions. Moreover, HDR productions will not be possible in every future situation. To adapt remaining SDR content to the future dominating HDR world, Inverse Tone Mapping is required, which creates HDR content by expanding the dynamic range of SDR material. While ordinary operations only focus on adjusting brightness values thus leading to visually unrealistic results, Deep-Learning-based approaches from the field of AI have recently led to qualitatively promising results. The paper will give an overview of restoring dynamic range especially by learning-based and data-driven approaches of Deep Learning and will provide technical fundamentals and examples of expanded images. Furthermore, the issue of adapting these approaches to live broadcast and post-production applications will be discussed, as compliance with technical requirements and quality standards is a challenging task in this respect.

INTRODUCTION

High Dynamic Range (HDR) has been widely adopted in several markets for quite some time and is already supported by the latest displays on the consumer market. Therefore, HDR has also become more and more important in the broadcast industry and will most likely become a production standard in the near future. However, most image content from almost 100 years of television is only available in lower “Standard” Dynamic Range (SDR) quality. This legacy material, e.g. from archives of television stations will always need to be included into current and future productions. Furthermore, due to economic limitations, the application of HDR may not be affordable for every production overnight. However, not only by lack of financial means but also technical restrictions can inhibit pure HDR productions, e.g. when being forced to use lower cost cameras with small non-HDR-capable sensors.

To adapt remaining SDR and restricted HDR content or even live images to future dominating HDR content and infrastructures as well as displaying it on new HDR devices, a so-called “Inverse Tone Mapping” (ITM) or “up-mapping” is required. Since the naive use of SDR content in HDR applications leads to very bad results, ITM operations create HDR content by expanding the contrast and dynamic range of SDR material. Ordinary operations, however, only focus on adjusting brightness values leading to visually unrealistic and poor results compared to original HDR images. Especially over- and underexposed image areas

with missing image information caused by technical or physical limitations of SDR camera sensors, are difficult to recover. Due to lost image information, restoring dynamic range is an ill-posed problem, which can be considered a problem of image recovery. This issue has also been researched in the field of artificial intelligence (AI) and has recently been adapted successfully to the HDR image reconstruction of still images leading to promising results. Lost brightness information can be estimated and reconstructed in a learning-based and data-driven manner using deep neural networks to mimic real HDR images. Thus, SDR material could be converted to HDR automatically, reducing cost and time.

The paper will give an overview of state-of-the-art Deep Learning approaches in the field of HDR image reconstruction and will provide technical fundamentals. Results will be shown using example imagery to verify performance and potential of this technique. Furthermore, the issue of adapting these approaches from the field of computer graphics to live broadcast and post-production applications will be discussed, since the use of Deep Learning approaches on moving image material while meeting the requirements of broadcast is a challenging task in this respect.

HIGH DYNAMIC RANGE IMAGING

Due to technical and physical limitations of SDR camera sensors, capturing in SDR leads to overexposed (saturated) and/or underexposed image regions, which in turn lead to missing image information. As a result, SDR leads to images that do not correspond to the real lighting conditions of a scene according to human visual perception. Unlike traditional SDR, HDR is able to capture, manipulate and display the real lighting conditions of a scene according to human perception. As shown in the comparison in Figure 1, the HDR image contains significantly more details in both shadows and (high-)lights. In contrast, the SDR image has lost image information in the bright overexposed areas of the image, e.g. in the sky or in reflections in the water, as well as in the darkest underexposed image regions. Moreover, HDR is not only capable of preserving these scenic details, but also of generating highlights at much higher brightness. However, this does not mean a simple increase in brightness over a large area, but rather a selective accentuation of the peak luminance values according to the lighting situation of the scene. These accentuations, which are usually caused by emissive light sources and specular reflections, are displayed noticeably more brilliant to be able to differentiate them from diffuse white (matte) surfaces. By providing lower black levels, a similar effect applies to shadow details. Altogether, HDR images appear much richer in contrast, more vivid and more natural thus covering the human visual perception of real scenes much better compared to traditional SDR images.

Figure 1 – Comparison: BT.709 SDR (left) and BT.2100 PQ HDR (right) [image: (2)]

To generate HDR images there are various approaches of High Dynamic Range Imaging (HDRI) that have been researched in the fields of computational photography and computer graphics (1). Among others, there are approaches extending the standard dynamic range of an image mathematically. Probably the most well-known approach of HDRI is to merge multiple SDR images captured with different exposures (e.g. auto bracket) to create HDR images. Photographers used to create their HDR images for a long time using this so called “time sequential multi-exposure technique”, also known as “stack-based method”. These methods estimate the non-linear camera response function of the differently exposed SDR images and subsequently linearize these multi-exposure images by applying the inverse response function. Afterwards, the multi-exposures are merged into a single HDR image containing all over- and underexposed image regions of these multi-exposures. Although it is possible to obtain images that are close to the lighting conditions of the real world, these methods have certain limitations. Due to the difficulty of capturing multi-exposures for a given scene, artifacts may occur. Particularly when changes occur during scene capture, due to moving objects, changes in lighting conditions or camera movements, this technique is very sensitive. As a result, ghosting or tearing artifacts may occur (1). Although several efforts have been made to overcome this issue, they are not considered reliable in each scenario, e.g. in case of moving image material with fast and unpredictable camera panning and object movements, as often the case in broadcast. Therefore, this method achieves good results in controlled scenes, but these rarely occur. Moreover, this approach is not directly suitable for HDR image reconstruction, as it cannot be applied to existing legacy material as this is usually not available in multiple exposures. Furthermore, such stack-based methods are lengthy and partly manual processes taking some time to be applied.

To overcome the disadvantage of stack-based methods, further approaches in HDRI have been developed (1). One of these approaches focus on extending the dynamic range by modifying the pixel architecture of ordinary camera sensors, e.g. by using pixel coding. Another approach based on more recent types of camera sensors, uses logarithmic response functions and 8-16-bit per pixel encoding. Several HDR cameras of this kind are commercially available, even from broadcast camera manufacturers, but cannot be used in every situation as described in the introduction. Among other limitations associated with these approaches, one of the major limitations related to the issue under discussion is the fact that these approaches cannot be applied to already existing legacy material.

INVERSE TONE MAPPING

Due to the facts described in the introduction, a link between traditional SDR and modern HDR imaging must be established, since displaying SDR images on HDR devices leads to unpleasant quantization artifacts as well as flat and unnatural results due to missing image information in over- and underexposed image regions. This loss of information in the original SDR image, e.g. caused by clipping, quantization, tone mapping, gamma correction or by mistakes during image acquisition, is very difficult to recover, especially in terms of strongly “corrupted” single exposures. Therefore, restoring dynamic range is an ill-posed problem which can be considered a problem of image recovery. In addition, not only the details in dark and bright image areas need to be restored, but also the shiny highlights caused by emissive light sources and specular reflections must be generated in higher peak luminance.

To overcome these challenges and create enhanced viewing experiences, “Inverse Tone Mapping” (ITM) is required, which adapts SDR content to future dominating HDR content and infrastructures to display it on new HDR devices. In general, a Tone Mapping (TM) describes an operation which reduces the dynamic range of an input image to adapt it to the

dynamic range of the display technology. In other words, a Tone Mapping Operator (TMO) converts HDR images into the SDR format, which is why the inverse case, i.e. the conversion of SDR material to HDR, is called *Inverse* Tone Mapping. Hence, the executing application known as “Inverse Tone Mapping Operator” (ITMO) creates HDR content by expanding the dynamic range of SDR image material. One of the first ITMOs of such kind was proposed by Banterle et al (3) in 2006. Just like HDRI, ITM originates from the field of computer graphics, but instead of dealing with the acquisition of pixel values according to the absolute or relative luminance of a scene, ITM is rather concerned with restoring these pixel values to achieve absolute or relative scene luminance. Furthermore, ITMOs can either be used to convert already tone-mapped content back to the HDR format, i.e. to undo already applied TMOs, or to expand the dynamic range of original SDR image material thus creating new HDR content that has not been existing before. This paper will focus on the latter case.

Typically, an ITMO includes operations such as linearization, bit-depth expansion, contrast or range expansion and, at best, the generation of highlights. Considering an appropriate bit-depth expansion is urgently required in the context of linearization and range expansion, since a simple linear expansion can lead to banding artifacts resulting in unwanted contours. However, restoring image details in highlights, such as emissive light sources and specular reflections as well as in deep shadows is one of the most important tasks in ITM. Due to the lack of image information in SDR content, this is also one of the biggest challenges to be solved. To overcome this challenge, various ITMOs have been proposed in recent years to be able to generate HDR images out of single exposed SDR images automatically.

Traditional Inverse Tone Mapping

Traditional ITMOs are either based on fixed static or specific parameterized functions to expand SDR images to HDR. These operators first linearize SDR images by applying a gamma correction or, if known, the inverse of the camera response function. Subsequently, the linearized pixel values, which are approximately proportional to the scene luminance, are expanded to the full dynamic range of an HDR display by using a range expansion while considering quantization and compression artifacts as well as basic viewer preferences (1). Most traditional ITMOs applying such an operation usually differ in the actual approach of range expansion as well as in accuracy of linearization. Furthermore, the majority of these ITMOs can usually be classified in two categories: global and local operators.

While global ITMOs (4) (5) apply the same operation to all pixels ensuring content to be evenly extended, local ITMOs (3) (6) (7) take local dependencies into consideration, i.e. the operation is applied depending on image content (e.g. depending on certain image regions). In contrast to global ITMOs, local operators rely on more complex operations, typically applying an analytical function in combination with an expansion function (or expand map). Nevertheless, the studies of Akyüz et al (4) and Masia et al (5), both of which include psychophysical evaluations, clearly show a viewer preference for the results of global ITMOs over complex local ones. These studies show that local ITMOs are not very successful in reconstructing missing image information. Instead, these are more prone to cause artifacts, since the brightness of overexposed image areas is simply increased by these operators, leading to unwanted contour artifacts. In addition, inappropriately set parameters can result in the dynamic range being incorrectly expanded for certain input images. Therefore, local ITMOs often fail to provide desirable results when it comes to expanding SDR content and simultaneously maximizing subjective quality. However, global ITMOs are not suitable for this task either, as they are also only capable of adjusting brightness information.

Although these approaches provide good results in case of high-quality input images with low amounts of lost image information, legacy material rarely meets these conditions. When dealing with content affected by artifacts and lost image information, such ITMOs encounter major difficulties resulting in visually unrealistic and poor HDR images of limited quality in terms of linearization and reconstruction of over- and underexposed image regions. This is because traditional ITMOs are not able to sufficiently compensate for lost image information according to the HDR requirements described earlier. Since these operators “derive” or rather adjust image information only through specific assumptions or individual heuristics, they are considered model-driven. This specific, individual approach makes them unsuitable for coping with the large amount of cross-content legacy material. Although traditional ITMOs enable the integration and manipulation of SDR content in HDR infrastructures and workflows to display it on HDR devices, the corresponding restrictions are significant.

This indicates that the use of AI might be necessary to meet the high requirements of this image recovery problem to create high-quality images with extended dynamic range.

Deep-Learning-Based Inverse Tone Mapping

Fundamentals

In recent years, AI has been increasingly employed in the fields of image processing such as image reconstruction. Convolutional neural networks (CNNs) from the area of Deep Learning (a sub-area of AI) have made a significant contribution to this. These CNNs often consist of dozens or hundreds of layers, which are designed to analyse image data for features and patterns independently to learn from these images. The basic layer construct of such CNNs for analysing image data mainly consists of three different layers; convolution, activation and downsampling layers repeatedly arranged in this order as shown in the upper left of Figure 2. By applying convolution filters of different sizes to SDR input image data, convolution layers activate features within this data thus generating so-called feature maps. These convolution filters (or layers) include adaptive parameters such as filter weights and biases which will be updated during training. To enable faster and more efficient training, detected features are passed on by the activation layer, which maintains positive values and reduces or even clips off negative values to zero. This ensures that primarily activated features are passed on to the next layer to keep them in focus. Afterwards, a downsampling layer provides a simplified output of the input data by performing a non-linear downsampling. However, this operation can also be performed by specific convolution filters. Subsequently, the downsampled image data is passed on to the next convolution layer starting another such “cycle”. The number of parameters to be learned is further reduced with each down-sampling layer, but at the same time the extracted features increase in their complexity. Therefore, a CNN usually starts by extracting simple features like edges and corners before more complex features like structures, shapes, textures, or other object elements will be extracted in deeper network layers. The deeper the training image data enters the network, the more complex the features and therefore the correlations are that a network can learn from this data. However, after the SDR image data has been passed through several downsampling layers, this data has been downsampled so many times that the image resolution is reduced to just a few pixels. To create images with full resolution at the output, downsampling must be undone in the following network layers, e.g. by using upsampling (or deconvolution) layers. This kind of network architecture shown in Figure 2 is called U-Net (8) or autoencoder architecture. These autoencoders often used by CNN-based ITMOs are designed to analyse SDR input image data using an encoder and output HDR image data using a decoder. To get back high quality, high resolution HDR images from downsampled

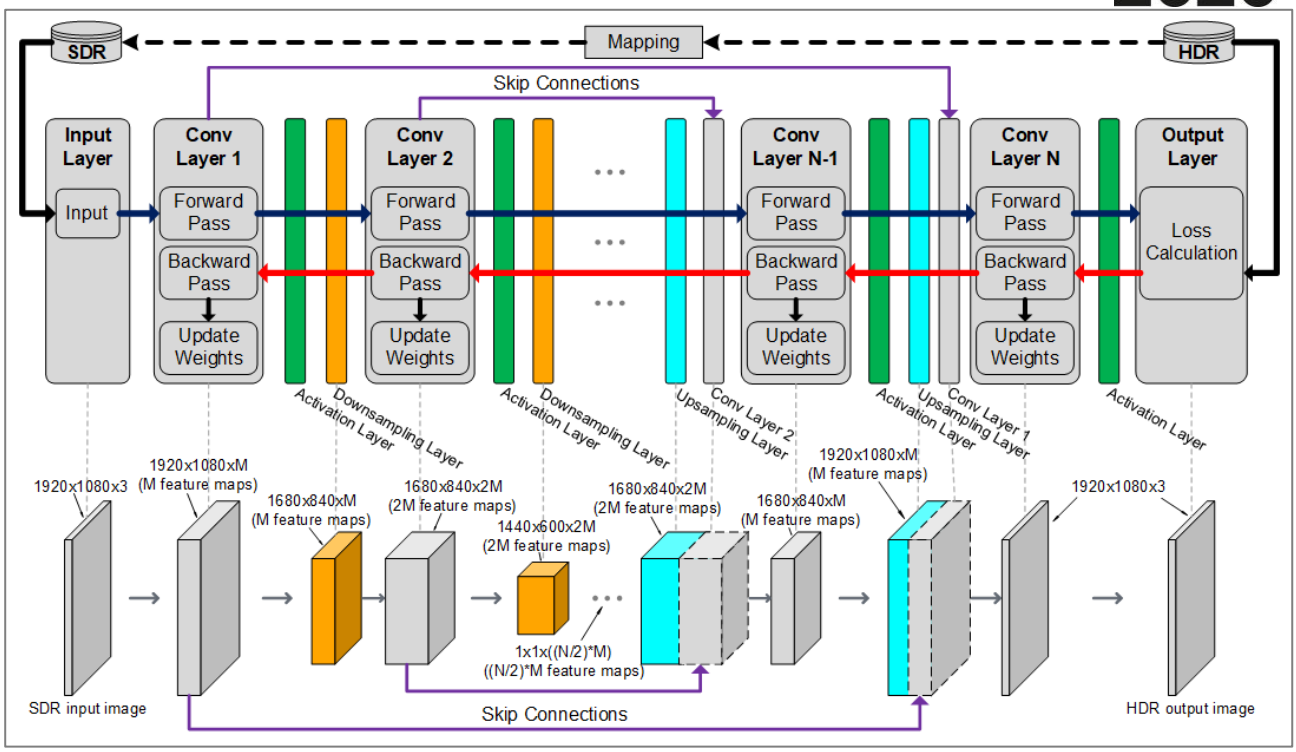

Figure 2 – Example scheme of a CNN-based autoencoder architecture

image data at the end of the network, these autoencoders make use of so-called skip-connections. These enable an exchange between lower layers of the encoder and deeper layers of the decoder (see Figure 2). Therefore, the decoder, which is responsible for upsampling, receives information on how corresponding image data looked like before it has been downsampled. This allows an upsampling operation to be “skipped”. Hence, skip-connections enable high-resolution image details to be fully exploit thus allowing deeper structures to be learned. After the SDR image data has passed through all network layers forward and HDR image data has been estimated at the end of the network, the loss between estimated data and the corresponding ground truth HDR data can be calculated by using a loss function in the output layer (see Figure 2). At this point the actual comparison between SDR and the corresponding HDR reference takes place. The loss function is basically an error function that compares the output of the network with the ground truth values. The greater the variance, the greater the loss or error. Therefore, the minimization of this error can be considered the global goal of such networks. By choosing an appropriate loss function, the network can effectively approach this goal. After the loss has been set off against the estimated data, the results are then backward propagated through the network. While each layer is passed through in reverse order, layers with learnable parameters (i.e. convolution layers) will then be updated by the calculated loss as shown in Figure 2. Therefore, filter weights and biases in these layers are optimized for their task of generating HDR images. After the network has been completely passed through by back-propagation and each layer has been updated, a new training iteration begins, in which SDR image data is passed through the forward pass of the network again. Now with updated parameters, the estimated image should get closer to the HDR reference than before. A new loss between the target image and the new estimated image can then be calculated and the parameters can be further optimized to minimize the error. Training is repeated in multiple iterations until the network can no longer learn from available training data and thus cannot further minimize the loss. The more training data is available, the more the network can learn.

Based on this technology, first approaches for HDR still image reconstruction have been proposed (9) (10) (11) (12) (13). These CNNs each accept single-exposed SDR images at the input and produce qualitatively promising HDR images at the output. Training of these networks is based on large and mostly diverse HDR image data sets, which are used to generate corresponding SDR images (see Figure 2) representing the conditions of legacy material. Therefore, reconstruction of lost image information due to quantization, clipping, TM or gamma correction can be learned by a high variety of features and correlations. However, CNN-based reconstruction of HDR images from single-exposed SDR images can be performed in two approaches: either indirectly or directly. Indirect approaches (9) (10) resolve the task via an intermediate step, in which several exposures are estimated from the SDR input image thus generating a multi-exposure image set, similar to the described HDRI approach of time sequential multi-exposure technique. According to this technique, the generated multi-exposures are subsequently linearized and merged into one HDR image. Direct approaches, on the other hand, generate HDR images directly from single-exposed SDR images, i.e. no intermediate step of creating multi-exposures is required. Since the sequential processing steps of estimating and linearizing multiple images require a higher processing time, indirect approaches are rather unsuitable for live applications. For this reason, this paper will focus on direct approaches.

CNN-based ITMOs

One of the first direct CNN-based ITMO was proposed by Eilertsen et al (11) in 2017. This ITMO specifically addresses the problem of estimating lost information in saturated image regions. After estimating the pixel values within these regions, they are subsequently combined with the linearized non-saturated regions. The linearization of these non-saturated regions is done beforehand by applying a fixed function without considering the actual response function. However, this ITMO is specialized in restoring overexposed image regions of rather underexposed SDR images. As a result, highly overexposed regions in SDR input images may not be sufficiently restored. Furthermore, this method is not capable of restoring underexposed image regions at all. The network is based on a hybrid dynamic range autoencoder, which makes use of skip-connections and works according to the U-Net architecture described above.

Furthermore, another direct CNN-based ITMO was proposed by Marnerides et al (12) in 2018, called “ExpandNet”. Unlike the CNN of Eilertsen et al (11), ExpandNet does not use straightforward autoencoder structures. Instead, the network uses an end-to-end multiscale architecture consisting of three parallel branches; a local branch, a dilation branch and a global branch. These branches perform their own specific tasks in parallel to each other respectively before the outputs are fused together. While the local branch learns how to maintain and extend high-frequency details on a local level, the dilation branch learns about larger pixel neighbourhoods. In addition, the global branch provides general information by learning the global context of the input image data. Moreover, the architecture of ExpandNet is designed to avoid upsampling layers. The global branch is the only one downsampling input image data, however, the output only gets replicated instead of upsampled when fusing the outputs together. Therefore, blocking and halo artifacts can be avoided, which may be caused by autoencoder networks. Moreover, by using this approach, ExpandNet can restore both overexposed and underexposed image regions.

Another direct CNN-based ITMO called “iTM-Net” was proposed by Kinoshita and Kiya (13) in 2019. By using an autoencoder structure and splitting the encoder into a local and a global encoder, Kinoshita and Kiya combine the U-Net architecture used by Eilertsen et al (11) with

the multiscale architecture used by Marnerides et al (12). The studies of Kinoshita and Kiya primarily point out that training CNN-based ITMOs using a standard loss function is a challenging task due to the non-linearity between SDR and HDR images. This is because HDR pixel values, unlike those in SDR, are non-uniformly distributed over an extremely wide range. By using a novel loss function during training and therefore considering this non-linear relationship, Kinoshita and Kiya were able to significantly improve the performance of their network. Instead of estimating HDR images directly, iTM-Net estimates tone-mapped versions of HDR images (i.e. SDR images). By applying a nonlinear but invertible global TMO to the ground truth HDR reference image, the distance between this properly tone mapped SDR image, corresponding to the relative dynamic range of the HDR reference, and the estimated SDR image can be considered within the loss calculation. Moreover, the global TMO ensures that the nonlinear relations between SDR and HDR pixel values are reduced when calculating the loss. Therefore, the loss function enables a distribution of the tone-mapped HDR pixel values according to the distribution of those in SDR. In case of employing the ITMO, SDR images estimated by iTM-Net are converted to HDR by applying the inverse of the global TMO. As mentioned earlier, such fixed static ITMOs achieve good results for SDR images that are not or barely affected by missing image information. Since the network is capable of restoring both overexposed and underexposed image regions, such a simple operator achieves good results.

Based on subjective and/or objective evaluations, all direct CNN-based ITMOs have proven significant quality improvements compared to traditional ITMOs. Figure 3 shows that ExpandNet by Marnerides et al (12) achieves good objective results (bottom left) using the SDR image from Figure 1 as input image (top left). In contrast to the result of a traditional global ITMO (top right), details in the sky and reflections in the water could be reconstructed to a certain extent by ExpandNet. The figure also shows that the result is quite close to the ground truth HDR (bottom right) in some respects. In addition, the evaluation by Marnerides

Figure 3 – BT.709 SDR (top left), traditional global ITMO (top right), Marnerides et al (12) + slight modification (bottom left), Ground Truth BT.2100 PQ HDR (bottom right) [image: (2)]

et al (12) shows an improvement in quality compared to the ITMO proposed by Eilertsen et al (11). Furthermore, Kinoshita and Kiya (13) demonstrate a further improvement compared to the network proposed by Marnerides et al (12).

Conclusion on CNN-based ITMOs

Due to the improvements in HDR image reconstruction, CNN-based ITMOs have attracted much attention in recent years. Especially due to the high performance and efficiency, this approach is considered very promising. Since these ITMOs derive various relations between SDR and HDR images, they are able to reconstruct physical brightness information instead of simply adjusting the image information existing. Therefore, lost image information in single-exposed SDR images can be reconstructed thus generating statistically plausible HDR images of high perceptible quality. Moreover, these ITMOs work fully automatically and data-driven, which is why no human expertise or heuristics are required (12), except for designing the network and selecting appropriate training data. These are the reasons why CNN-based ITMOs offer enormous potential when it comes to converting the large amounts of diverse legacy content to HDR thus creating enhanced viewing experience. Since the ITMO by Eilertsen et al (11) is affected by too many limitations such as underestimation of highlights, difficulties with highly overexposed image regions and the inability to restore underexposed regions at all, special focus should be given to the approaches of Marnerides et al (12) and especially Kinoshita and Kiya (13). However, these approaches still need to be further improved to produce high-quality HDR images while meeting the requirements of live broadcast.

GAN-based ITMOs

In addition to the CNN-based ITMOs, there is another very recent Deep Learning approach (14) using generative adversarial networks (GANs) (15) instead of CNNs. GANs are based on a whole different approach consisting of two networks being trained simultaneously. One of these networks acts as a generator whose task is to capture the data distribution of input image data to generate samples. The second one acts as a discriminator estimating the probability that the sample came from the training data instead of being generated by the generator. The generator is played off against the discriminator by trying to maximize the probability of the discriminator making a mistake. During the learning process, the distribution generated by the generator is approximated to the real distribution. The discriminator is also being improved during the learning process, enabling the generator to further approximate the generated samples to the real ones. Therefore, the competition between generator and discriminator pushes both networks to further improve their tasks until the generated result can no longer be distinguished from the original. Kim et al (14) proposed an ITMO based on GAN structure, called “JSI-GAN”. This network additionally combines ITM functionality with super-resolution, a learning-based approach for upscaling image resolution. For ITM functionality, Kim et al designed a subnet for image reconstruction, which is complemented by two further subnets for local contrast enhancement and detail restoration, respectively. Kim et al prove that JSI-GAN achieves results of higher quality compared to traditional ITMOs. Furthermore, improved quality compared to the CNN-based approach of Eilertsen et al (11) is demonstrated. Since GANs were initially intended to produce photorealistic images, these networks could be well suited for HDR image recovery tasks.

DISCUSSION

Why Is Deep-Learning-Based ITM Relevant for Broadcast At All?

As already described in the introduction, both the large amounts of existing legacy material and the SDR material captured in the future will always need to be converted into the HDR format to enable pure HDR productions. In addition, viewers should not be allowed to notice any difference between SDR and HDR images. Instead, viewing experience should fully benefit from more realistic and vibrant images regardless of whether content is or has been captured in SDR or HDR. Consequently, a link between traditional SDRI and modern HDRI must be established since displaying SDR images on HDR monitors leads to unpleasant quantization artifacts and flat, unnatural results due to missing image information in over- and underexposed image regions. Moreover, it is not only the details in dark and bright areas of the image that need to be restored, but also the shiny highlights caused by emissive light sources and specular reflections that must be generated in higher peak luminance. To overcome these challenges and create enhanced viewing experience, adapting ITM approaches from the field of computer graphics to broadcast applications is necessary. The approaches and studies presented in this paper clearly show that learning-based ITMOs are much better suited to accomplish the task of HDR image reconstruction compared to traditional ITMOs. Instead of using fixed static or specific parameterized functions which are not suitable for the large amounts of diverse legacy material, learning-based approaches derive various dependencies and relations between HDR and SDR image data by using large training data sets. In contrast to traditional ITMOs based on model-driven approaches and therefore extending the dynamic range only by adjusting image information, learning-based and data-driven ITMOs extend the dynamic range by simultaneously reconstructing real physical brightness information. Therefore, structures and shapes in critical image regions can be restored to a certain extent and details in highlights and shadows can be reconstructed to mimic real HDR images. Moreover, learning-based approaches have proven to be less prone to generating artifacts thus leading to visually promising results of high perceptible quality. Furthermore, if such approaches can be adapted to broadcast applications, SDR material could be converted into the HDR format fully automatically. Therefore, additional time-consuming and manual steps for restoring SDR images during (post-)production could be avoided, thus reducing time and cost.

Challenges in Adapting ITM Approaches to (Live) Broadcast

Due to high technical requirements and quality standards in (live) broadcast and the high complexity of learning-based ITMOs, combining both worlds is a challenging task. One of the most important requirements in broadcast is to create realistic and natural high-quality HDR images matching the viewer's preference. Hence, the resulting look should not strongly differ from the original HDR look and generated content should be free of any quality degrading artifacts. Although learning-based ITMOs have proven to be on a good path to meeting these requirements for still image reconstruction, in case of moving image material the resulting image quality of these ITMOs still needs to be researched. If necessary, adjustments or precautions must be made to avoid potential artifacts such as contouring, banding, ghosting, tearing, flickering or halos, especially caused by unforeseeable changes in moving images. This includes program changes, scene cuts, changes in lighting conditions and changes caused by object or camera movements, especially when occurring fast and unpredictably. As these unforeseen changes often occur in live broadcast situations, special attention should be paid to them to strictly avoid artifacts at any time.

Therefore, systems performing such image processing should run very stable. In addition, the results of existing approaches need to be further improved. In this context, not only the results in overexposed, but especially in underexposed image regions must be optimized to meet the broadcast quality requirements in the best possible way. However, probably the most important requirement for image processing systems in live broadcast is to meet real-time capability which is essential for live broadcast applications. Due to their complexity, learning-based ITMOs have not been able to meet this requirement yet, which is probably one of the most challenging hurdles to overcome. Although some of these approaches can process Full HD images, they require too much processing time even on modern hardware. Hence, special software or hardware solutions are required to enable real-time processing. As a result, the optimization of learning-based ITMOs to moving images and the fulfilment of real-time requirements while providing high image quality are the main challenges to be overcome when adapting learning-based approaches to (live) broadcast applications.

CONCLUSION AND OUTLOOK

Due to various reasons discussed, converting SDR image content to the HDR world will be essential in future broadcast. However, HDRI approaches are not applicable to existing SDR legacy material and cannot be used for every (live) broadcast situation. For this reason, and because the naive use of SDR content in HDR applications is far from sufficient, Inverse Tone Mapping is required. Various approaches and studies clearly show that traditional operators are not suitable for converting the large amounts of cross-content legacy material appropriately due to their specific model-driven approaches. Moreover, these operators are only able to adjust the available brightness information, so most of them are prone to cause artifacts when image information is missing. Instead, the approaches and evaluations presented clearly show that learning-based operators must be focused to overcome this ill-posed problem of HDR image reconstruction. Due to their data-driven approach, these are well suited to cope with the large amounts of diverse legacy material. In contrast to traditional operators, learning-based approaches can reconstruct physical brightness information making them less prone to cause artifacts. Therefore, such operators produce high quality results with natural tones and less visible noise even in critical image regions. Furthermore, technical fundamentals of learning-based approaches, especially relevant network architectures such as autoencoders and multiscale structures have been described. Moreover, the significance of appropriate loss calculation has been stressed and described in more detail. Nevertheless, learning-based approaches do not yet meet the broadcast requirements, particularly in terms of real-time capability and moving image material, both of which are considered very demanding. In addition, the application to moving images may be associated with potential artifacts affecting image quality. Hence, these approaches still need to be further improved, e.g. by optimising and adapting network structures and parameters. For this purpose, related learning-based approaches addressing similar image recovery problems, e.g. from the fields of denoising, bit-depth expansion and image inpainting could be consulted. With GANs, another very different network type was described, which is also considered promising for HDR image reconstruction tasks. These networks should be focused and studied more closely in the future to further improve HDR image quality generated by learning-based ITMOs.

REFERENCES

1. Mantiuk, R. K., Myszkowski, K. and Seidel H.-P., 2015. High Dynamic Range Imaging. Wiley Encyclopedia of Electrical and Electronics Engineering. June, 2015.

2. Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A. and Brendel, H., 2014. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR displays. Proc. SPIE. vol. 9023.

3. Banterle, F., Ledda, P., Debattista, K. and Chalmers, A., 2006. Inverse Tone Mapping. GRAPHITE '06. November, 2006. pp. 349 to 356.

4. Akyüz, A. O., Fleming, R., Riecke, B. E., Reinhard, E. and Bülthoff, H. H., 2007. Do HDR Displays Support LDR Content? A Psychophysical Evaluation. ACM Trans. Graph. vol. 26. pp. 38-es.

5. Masia, B., Agustin, S., Fleming, R. W., Sorkine, O. and Gutierrez, D., 2009. Evaluation of Reverse Tone Mapping Through Varying Exposure Conditions. ACM Trans. Graph. vol. 28. pp. 1 to 8.

6. Rempel, A., G., Trentacoste, M., Seetzen, H., Young, H. D., Heidrich, W., Whitehead, L. and Ward, G., 2007. Ldr2Hdr: On-the-fly Reverse Tone Mapping of Legacy Video and Photographs. ACM Trans. Graph. vol. 26. article 39.

7. Kovaleski, R. P. and Oliveira, M. M., 2014. High-Quality Reverse Tone Mapping for a Wide Range of Exposures. SIBGRAPI 2014. pp. 49 to 56.

8. Ronneberger, O., Fischer, P. and Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015. vol. 9351. pp. 234 to 241.

9. Endo, Y., Kanamori, Y. and Mitani, J., 2017. Deep Reverse Tone Mapping. ACM Trans. Graph. vol. 36. article 177.

10. Lee, S., An, G. H. and Kang, S.-J., 2018. Deep Chain HDRI: Reconstructing a High Dynamic Range Image from a Single Low Dynamic Range Image. IEEE Access. vol. 6. pp. 49913 to 49924.

11. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R. K. and Unger, J., 2017. HDR image reconstruction from a single exposure using deep CNNs. ACM Trans. Graph. vol. 36. article 178.

12. Marnerides, D., Bashford-Rogers, T., Hatchett, J. and Debattista, K., 2018. ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. Computer Graphics Forum. vol. 37. pp. 37 to 49.

13. Kinoshita, Y. and Kiya, H., 2019. iTM-Net: Deep Inverse Tone Mapping Using Novel Loss Function Considering Tone Mapping Operator. IEEE Access. vol. 7. pp. 73555 to 73563.

14. Kim, S. Y., Oh, J. and Kim, M., 2020. JSI-GAN: GAN-Based Joint Super-Resolution and Inverse Tone-Mapping with Pixel-Wise Task-Specific Filters for UHD HDR Video. AAAI 2020. February, 2020.

15. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative Adversarial Nets. NIPS'14. vol. 2. pp. 2672 to 2680.

ACKNOWLEDGEMENTS

The author would like to thank his colleagues, fellow students and professors for their extremely valuable contributions to this work. He would also like to thank the International Broadcasting Convention for permission and support to publish this paper.