

COLORIZING SENTINEL-1 SAR IMAGES USING A VARIATIONAL AUTOENCODER CONDITIONED ON SENTINEL-2 IMAGERY

M. Schmitt¹, L. H. Hughes¹, M. Körner², X. X. Zhu^{1,3}

¹ Signal Processing in Earth Observation, Technical University of Munich (TUM), Munich, Germany - (m.schmitt,lloyd.hughes)@tum.de

² Chair of Remote Sensing Technology, Technical University of Munich (TUM), Munich, Germany - marco.koerner@tum.de

³ Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany - xiao.zhu@dlr.de

Commission II, WG II/6

KEY WORDS: Synthetic aperture radar (SAR), optical remote sensing, Sentinel-1, Sentinel-2, deep learning, data fusion

1. INTRODUCTION

Synthetic aperture radar (SAR) images are completely different from optical images in terms of both geometric and radiometric appearance: While SAR is a range-based imaging modality and measures physical properties of the observed scene, optical imagery basically represents an angular measurement system and collects information about the chemical characteristics of the environment. Thus, the interpretation of SAR imagery is still a challenging task for remote sensing scientists. However, SAR image interpretation can be alleviated when optical colors are used to support the interpretation process. For decades, this has been a special case of remote sensing image fusion (Pohl and van Genderen, 1998; Schmitt and Zhu, 2016). Still, SAR-optical image fusion by definition needs both a SAR and an optical image acquired at approximately the same time, which means standard image fusion techniques do not particularly help the interpretability of SAR images as an independent data source. To overcome the need for accompanying optical imagery, this paper proposes to learn feasible colorizations of Sentinel-1 SAR images from co-registered Sentinel-2 training examples using deep learning techniques. This is meant to provide a significant step in SAR-optical data fusion (Schmitt et al., 2017) with application to improved SAR image understanding, and will enable SAR data providers to attach colorized versions of their imagery to their products.

In order to achieve this goal, we find inspiration in recent computer vision approaches dealing with gray-scale image colorization. In the frame of this paper, we use the network architecture proposed by Deshpande et al. (2017), which utilizes both a *variational autoencoder (VAE)* as well as a *mixture density network (MDN)* (Bishop, 1994) to create multi-modal colorization hypotheses. Since our problem is different from the computer vision task in that there are no color SAR images that can be used as target samples during training, we first create artificial color SAR images by SAR-optical image fusion. In the remainder of this paper, we will first describe how these artificial color SAR images can be created (Section 2). Then, we introduce our dataset based on Sentinel-1 and Sentinel-2 imagery in Section 3, before we describe the employed deep generative model in Section 4. Finally, exemplary results are shown in Section 5.

2. SAR-OPTICAL IMAGE FUSION BY COLOR SPACE TRANSFORM

As Pohl and van Genderen (1998) have already summarized two decades ago, image fusion by exploitation of color space transformation has long been an established approach in remote sensing.

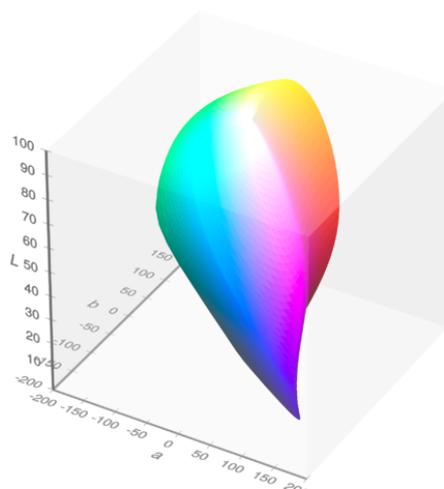


Figure 1. The visible gamut within the *Lab* color space.
 ©Wikimedia Commons / CC-BY-SA-3.0

In particular, the so-called *intensity-hue-saturation (IHS)* transform has become a standard procedure for SAR-optical image fusion (Harris and Murray, 1990), making use of the advantage over the conventional RGB color space that intensity or brightness is disentangled from the image colors (Tu et al., 2001). However, we follow a similar fusion strategy that employs a transformation of the optical imagery to *Lab* color space, which has the advantage that it describes all perceivable colors in a three-dimensional Cartesian coordinate system which reserves two dimensions for the actual color components and an additional luminosity axis (see Fig. 1). It has to be noted that the usage of the term *Lab* is not clearly defined and commonly used for different color space variants. In the context of this paper, we use it to refer to the *CIE 1976 L*a*b** color space (Pauli, 1976).

For our case of SAR-optical image fusion, the *Lab* color transform can be exploited in the following way: First, the optical image I_{opt} comprised of RGB pixels $\mathbf{p}_{\text{RGB}} = (R, G, B)^T$ is converted to *Lab* color space by determining the luminosity

$$L = 116 \cdot \sqrt[3]{\frac{Y}{Y_n}} - 16, \quad (1)$$

the value on the green-red color axis

$$a = 500 \cdot \left(\sqrt[3]{\frac{X}{X_n}} - \sqrt[3]{\frac{Y}{Y_n}} \right), \quad (2)$$

and the value on the yellow-blue color axis

$$b = 200 \cdot \left(\sqrt[3]{\frac{Y}{Y_n}} - \sqrt[3]{\frac{Z}{Z_n}} \right), \quad (3)$$

where the the intermediate XYZ color space representation

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (4)$$

is calculated by linearly mapping from RGB space and $X_n = 95.047$, $Y_n = 100$, $Z_n = 108.883$ are defined by the CIE standard illuminant D65 (Noboru and Robertson, 2005). Afterwards, the optical luminosity L is replaced by a SAR-derived pseudo-luminosity

$$L_{\text{SAR}} = 100 \cdot \sigma^0, \quad (5)$$

where σ^0 is the terrain-corrected SAR backscatter coefficient in decibel scale and the factor 100 is needed to bring $\sigma^0 \in [0; 1]$ to the range of luminosity $L \in [0; 100]$. Then the new *Lab* color triplet $\mathbf{p}_{\text{Lab}} = (L_{\text{SAR}}, a, b)^T$ is transformed back to RGB color space to obtain the fused image \mathbf{I}_{Lab} . Depending on the preparation of the SAR-derived luminosity, \mathbf{I}_{Lab} is similar to the example shown in Fig. 2. We use such Lab-fusion-based artificial color SAR images as targets in the training of the colorization model described in Section 4.

3. THE DATASET

The experiments presented in this paper make use of a dataset consisting of co-registered SAR and optical image patches sized 256×256 px with a pixel spacing of 10 m. The patches were cut from hundreds of georeferenced Sentinel-1 and Sentinel-2 images downloaded from uniformly distributed regions of interest spread across the land masses of the Earth. For the Sentinel-1 SAR images, GRD products were used, which represent the σ^0 backscatter coefficient in dB scale. Restituted orbit information was combined with the 30 m-SRTM-DEM (or the ASTER DEM for high latitude regions where SRTM is not available) to produce precisely ortho-rectified images. While the original data is quantized to 16 bit, the images used in this study were reduced to 8 bit. For Sentinel-2, the red, green, and blue channels (i.e. bands 4, 3, and 2) corresponding to the same areas of interest were downloaded and reduced to 8 bit as well. Since Sentinel-2 data are not provided in the form of satellite images, but as precisely georeferenced *granules*, no further pre-processing was required. Instead, for Sentinel-2, the data were selected based on the amount of cloud coverage. In order to not discard too many images, during the database query less than or equal to 1% of cloud coverage was used as search criterion.

Since for a given, randomly selected, region of interest, full coverage by single Sentinel-1/Sentinel-2 acquisitions cannot be ensured, mosaicking was used, whereas only images from within a meteorological season (i.e. spring, summer, fall, winter) were combined. However, automatic mosaicking of Sentinel-1 imagery can lead to artifacts such as seam lines. Furthermore, some clouds can still be present in the Sentinel-2 imagery. Therefore, all the produced image patches were manually inspected in order to remove affected patches. This way, in total 282,384 quality-controlled patch-pairs covering all continents and all seasons were produced.



(a)



(b)



(c)

Figure 2. Example for *Lab*-based image fusion: (a) logarithmic Sentinel-1 image \mathbf{I}_{SAR} , (b) Sentinel-2 RGB image (bands 4, 3, 2) \mathbf{I}_{opt} , (c) fusion result \mathbf{I}_{Lab} . The fused image combines a SAR-derived luminosity with the colors of its optical counterpart.

In order to produce colorized SAR images as training examples, *Lab*-based image fusion, as described in Section 2, was carried out for each patch-pair.

4. LEARNING COLORIZATION

4.1 Deep Generative Architecture

In order to learn an automatic colorization model for SAR images \mathbf{I}_{SAR} , which does not need an accompanying optical image at production time, we make use of the *variational autoencoder (VAE)* and *mixture density network (MDN)* architecture developed by Deshpande et al. (2017). It aims at training a conditional color distribution from which different colorization hypotheses can be drawn. The basic design of this deep generative model is depicted in Fig. 3. Within this architecture, the VAE is employed to generate a low level embedding \mathbf{z} of the color field of the *Lab*-based colorized SAR image \mathbf{I}_{Lab} . Relying on this low-dimensional latent variable embedding, the MDN is used to generate a multi-modal conditional distribution $p(\mathbf{z}|\mathbf{I}_{\text{SAR}})$ that models the relationship between the gray-level SAR image \mathbf{I}_{SAR} and the color

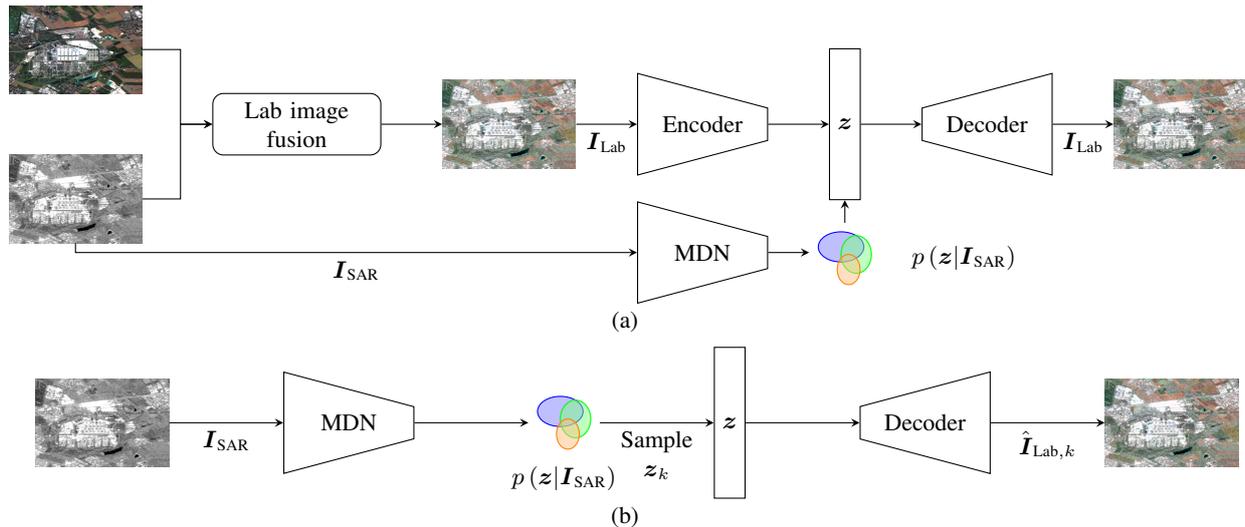


Figure 3. The proposed colorization network. (a) The training network architecture which uses both Sentinel-1 and Sentinel-2 data in order to learn feasible colorizations for SAR images. (b) The colorization test network which makes use of only a Sentinel-1 image in order to estimate a color distribution which is then sampled to colorize the original image.

embedding z . By sampling $z_k \sim p(z|I_{SAR})$ from this distribution, the decoder network of the VAE is able to generate a diverse range of colorized representations $\hat{I}_{Lab,k}$ of the original SAR image.

4.2 Implementation Details

For the VAE, which comprises the upper right part of Subfigure 3a, we chose to rely on the encoder-decoder architecture with skip connections that has provided the best results in (Deshpande et al., 2017). As for the losses, we also follow the optimal case, i.e. we use the full decoder loss as a sum of color rebalancing loss \mathcal{L}_{hist} , the specificity loss \mathcal{L}_{mah} , and the regularizing loss \mathcal{L}_{grad} .

For the MDN training, as suggested by Deshpande et al. (2017), we do not feed the SAR image I_{SAR} directly into the MDN, but rely on the features extracted up to the *conv7* layer of the colorization network proposed by Zhang et al. (2016), which was pre-trained on the 1.3 million images from the ImageNet training set (Russakovsky et al., 2015).

4.3 Training

We trained the network depicted in Subfigure 3a in a two-stage manner. Firstly, the VAE network was trained on 252,384 *Lab*-based fusion images using standard Adam optimization (Kingma and Ba, 2015) implementation for 15 epochs. The optimization hyper-parameters are fixed to $\beta_1 = 0.9$, $\beta_2 = 0.999$ with a learning rate of $\alpha_t = 5 \cdot 10^{-5}$.

The second stage was to train the MDN network. Training was performed using the corresponding 252,384 *conv7* gray-level features, as well as the latent codes produced by trained VAE for each of the *Lab*-based fusion training images. The MDN network was trained for 7 epochs, with the same β parameters as the VAE network and a learning rate of $\alpha_t = 10^{-3}$.

The β parameters are the default parameters recommended for the Adam optimization algorithm, while the learning rates were based on the details provided by Deshpande et al. (2017).

5. COLORIZATION RESULTS

For evaluation of the colorization capabilities of the architecture described in Section 4, we order the MDN-predicted gaussian mixture model means μ_i in descending order based on the mixture weights π_i and then display the results of the top-8 means for some example images of our test set comprising 1024 images. Some colorization examples drawn from the test set are shown in Fig. 4. It can be seen that Subfigures 4a to 4d can be considered as successful colorizations, as the target *Lab* appearance was well met, whereas the color adds valuable information to the SAR image, which supports general interpretability. For example, in results 4a, the rectangular field-like structures could possibly also have been caused by a fish farm or desalination ponds, while the optical color information rather suggests we are looking at something like rice fields. Another interesting phenomenon to be noted is the correct prediction of blue roofs in example 4d. While blue roofs are not uncommon in parts of China, they are rarely seen in the rest of the world. The colorization network, however, has successfully learned to predict them where applicable.

In contrast to these promising results, Subfigures 4e and 4f clearly belong to the class of failed examples. In these cases, the network does not only fail to match the target *Lab* color distribution, but also does not add any valuable information that would help image interpretation endeavours. Probably caused by the fact that many natural surfaces show a green or brown color appearance, the colors are predicted close to the corresponding expectation value.

The results indicate that colorizing SAR images based on a combination of *Lab*-based SAR-optical image fusion and a deep generative neural network architecture is able to produce artificial color SAR images, which can greatly support human operators in the task of image interpretation.

6. SUMMARY AND CONCLUSION

In this paper, we have shown an approach for the automatic colorization of SAR backscatter images, which are usually provided in the form of single-channel gray-scale imagery. Using a deep

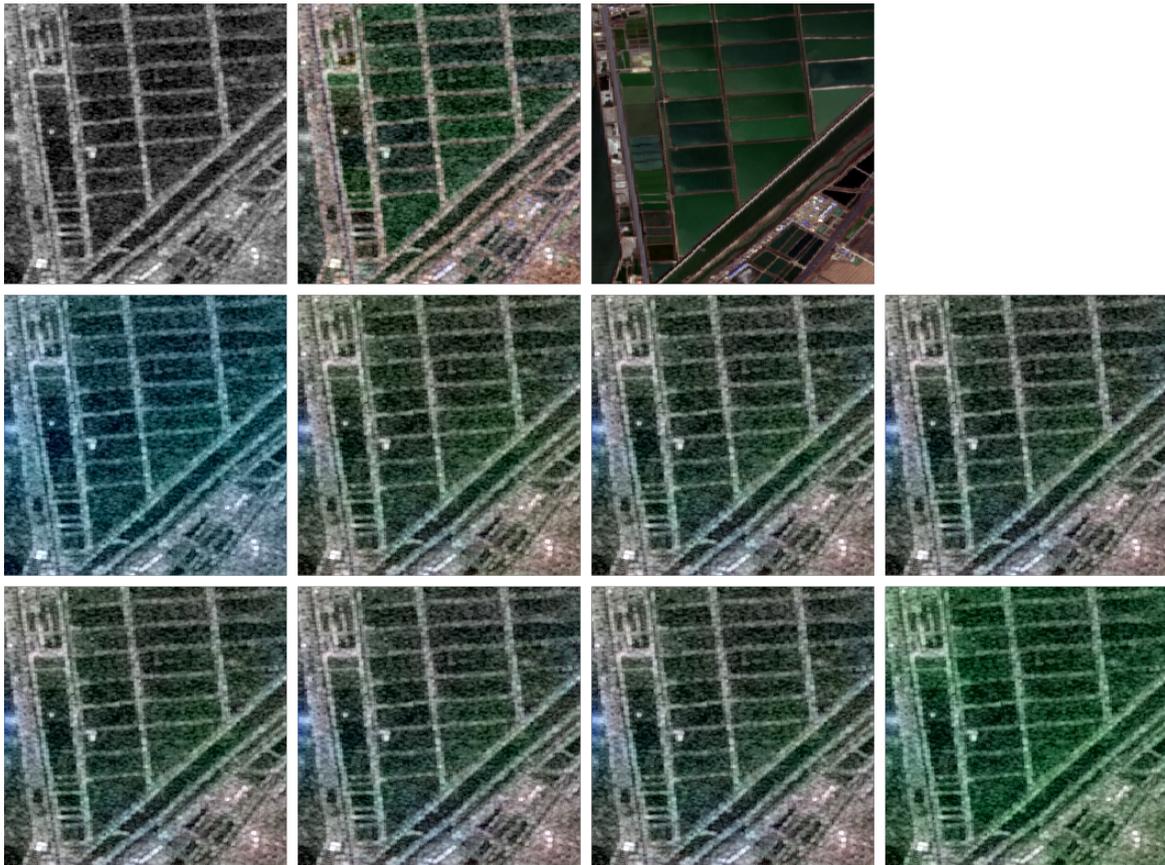
generative model proposed for the purpose of photograph colorization and a *Lab*-space-based SAR-optical image fusion formulation, we are able to predict artificial color SAR images, which disclose much more information to the human interpreter than the original SAR data. Future work will aim at further adaption of the employed procedure to our special case of multi-sensor remote sensing imagery. Furthermore, we will investigate if the low-level representations learned intrinsically by the deep network can be used for SAR image interpretation in an end-to-end manner.

ACKNOWLEDGEMENTS

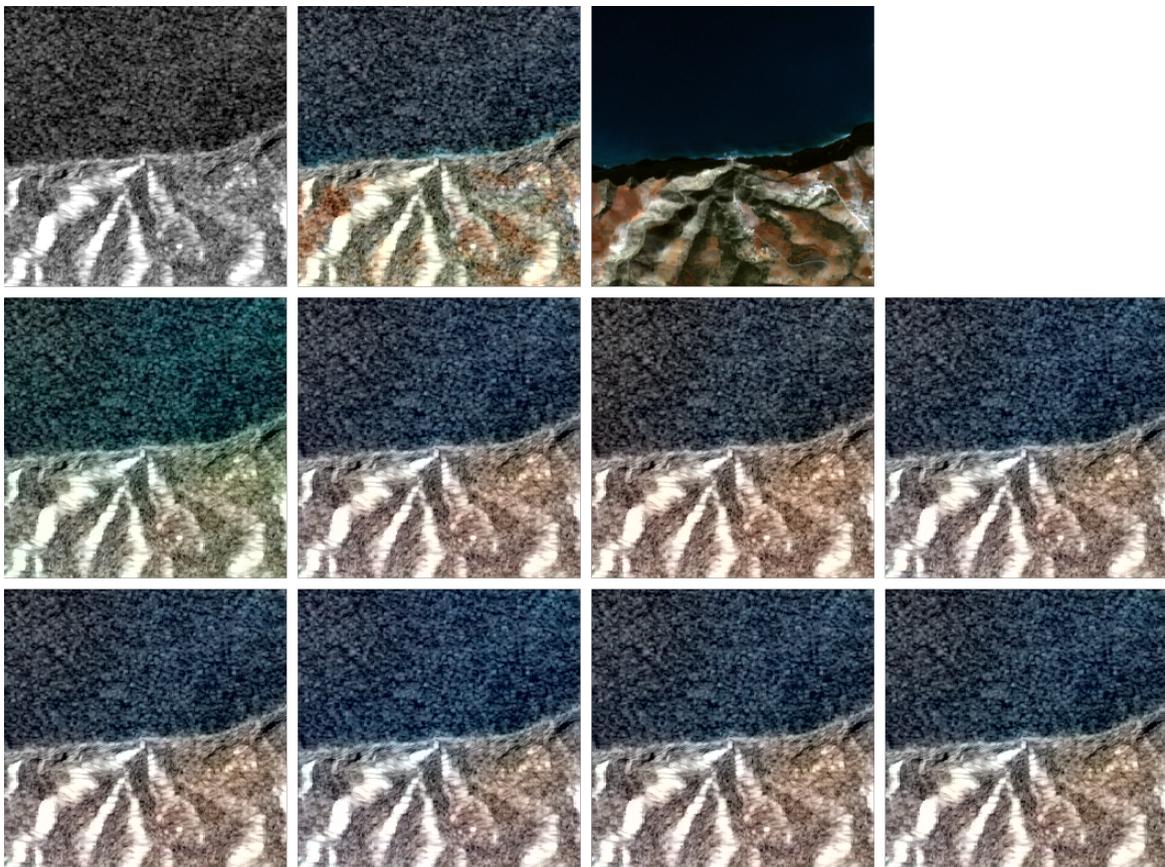
This work is jointly supported by the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018, www.sipeo.bgu.tum.de), the German Research Foundation (DFG) under grant SCHM 3322/1-1, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. ERC-2016-StG-714087, Acronym: *So2Sat*).

References

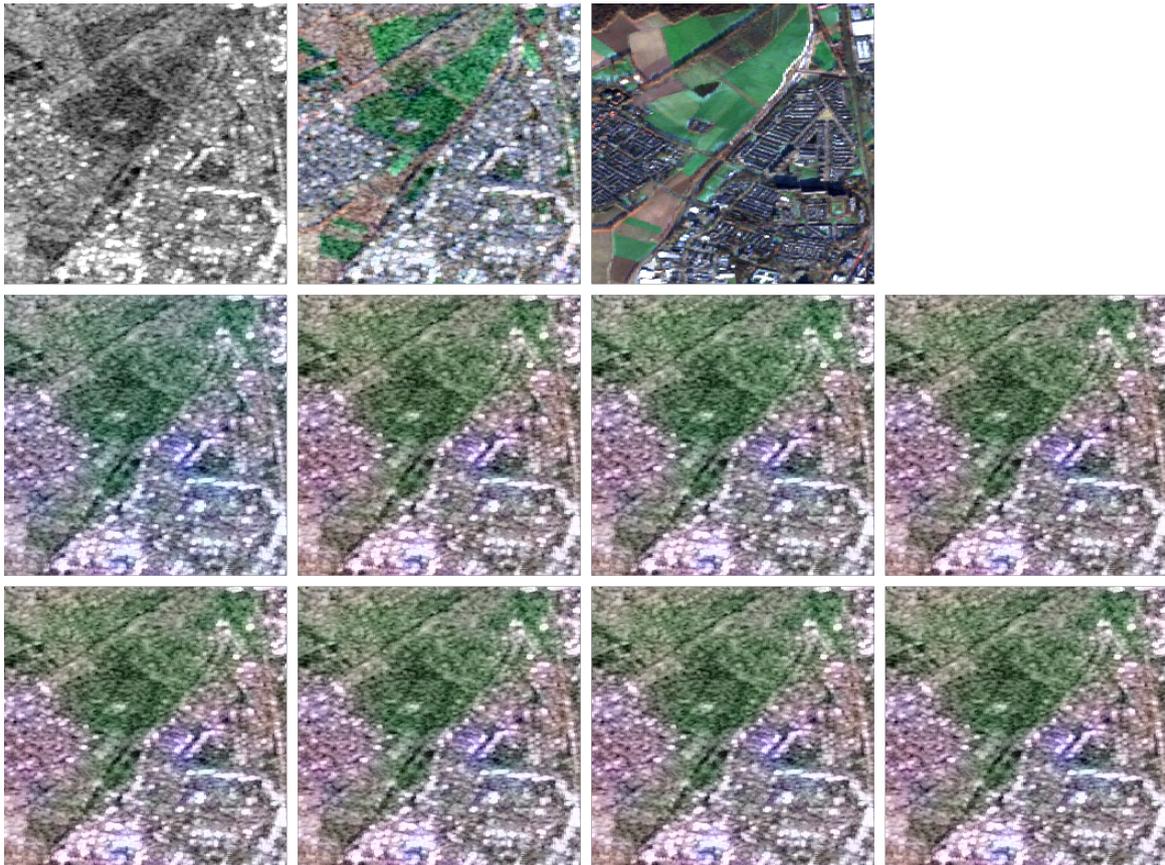
- Bishop, C. M., 1994. Mixture density networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University.
- Deshpande, A., Lu, J., Yeh, M.-C., Chong, M. J. and Forsyth, D., 2017. Learning diverse image colorization. In: *Proc. CVPR*, Honolulu, HI, USA, pp. 6837–6845.
- Harris, J. R. and Murray, R., 1990. IHS transform for the integration of radar imagery with other remotely sensed data. *56*(12), pp. 1631–1641.
- Kingma, D. and Ba, J., 2015. Adam: A method for stochastic optimization. In: *Proc. ICLR, arXiv:1412.6980v9*, San Diego, U.S.A.
- Noboru, O. and Robertson, A. R., 2005. Standard and supplementary illuminants. In: *Colorimetry*, Wiley, pp. 92–96.
- Pauli, H., 1976. Proposed extension of the CIE recommendation on "Uniform color spaces, color difference equations, and metric color terms". *Journal of the Optical Society of America* *66*(8), pp. 866–867.
- Pohl, C. and van Genderen, J. L., 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing* *19*(5), pp. 823–854.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* *115*(3), pp. 211–252.
- Schmitt, M. and Zhu, X., 2016. Data fusion and remote sensing – an ever-growing relationship. *IEEE Geosci. Remote Sens. Mag.* *4*(4), pp. 6–23.
- Schmitt, M., Tupin, F. and Zhu, X., 2017. Fusion of SAR and optical remote sensing data - challenges and recent trends. In: *Proc. IGARSS*, Fort Worth, TX, USA, pp. 5458–5461.
- Tu, T.-M., Su, S.-C., Shyu, H.-C. and Huang, P. S., 2001. A new look at IHS-like image fusion methods. *Information Fusion* *2*(3), pp. 177–186.
- Zhang, R., Isola, P. and Efros, A., 2016. Colorful image colorization. In: *Proc. ECCV*, Amsterdam, The Netherlands, p. 649666.



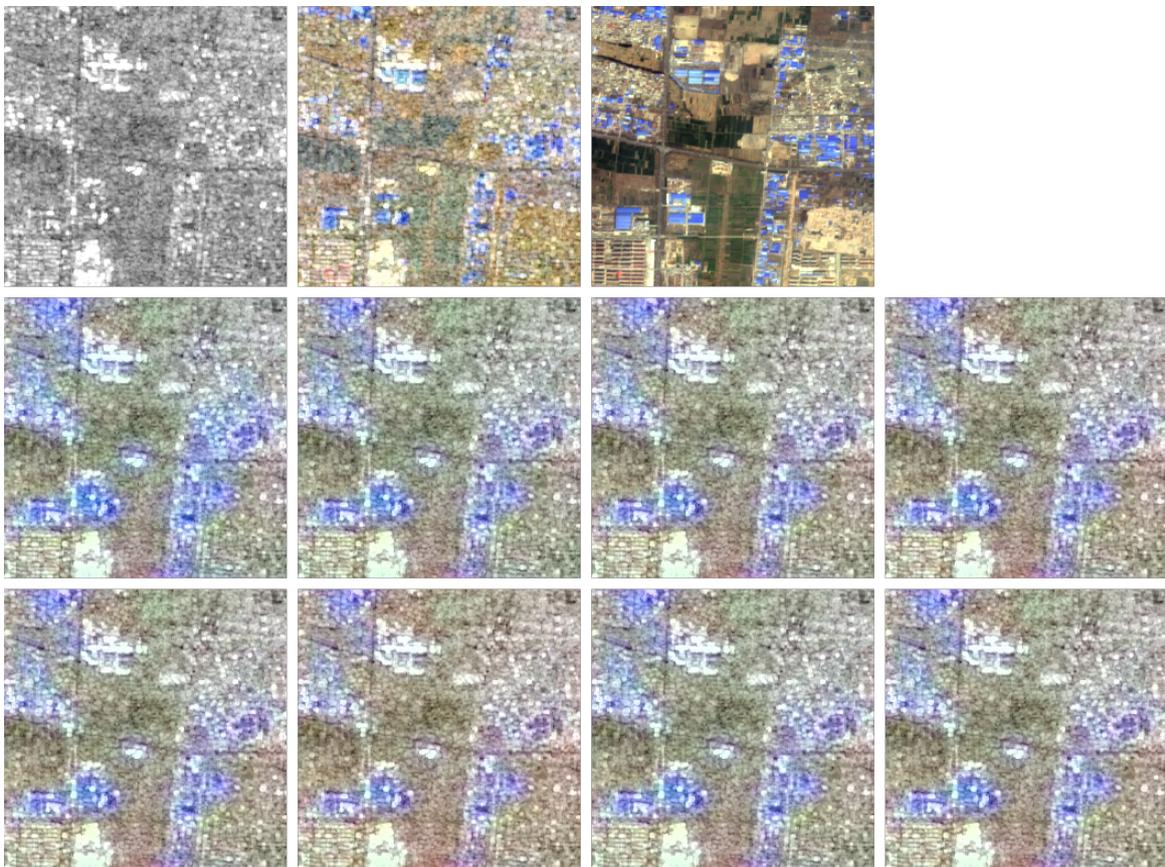
(a)



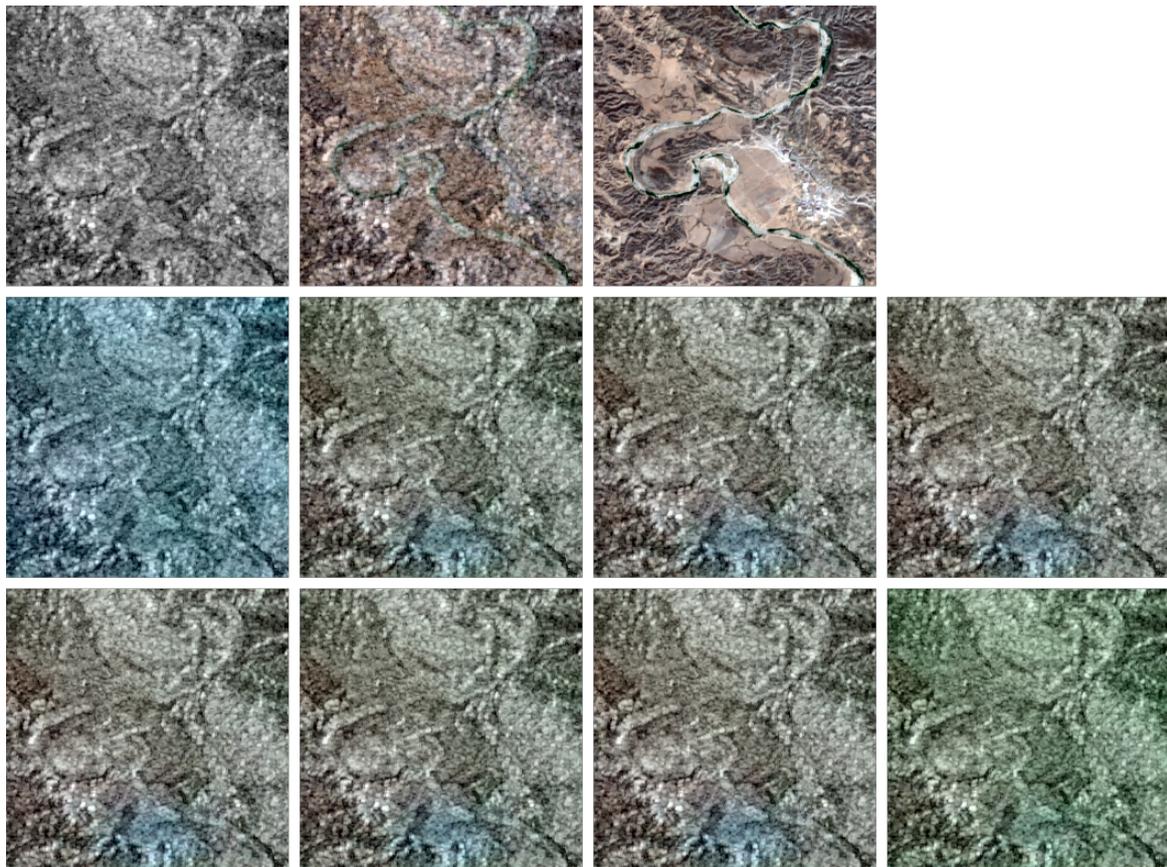
(b)



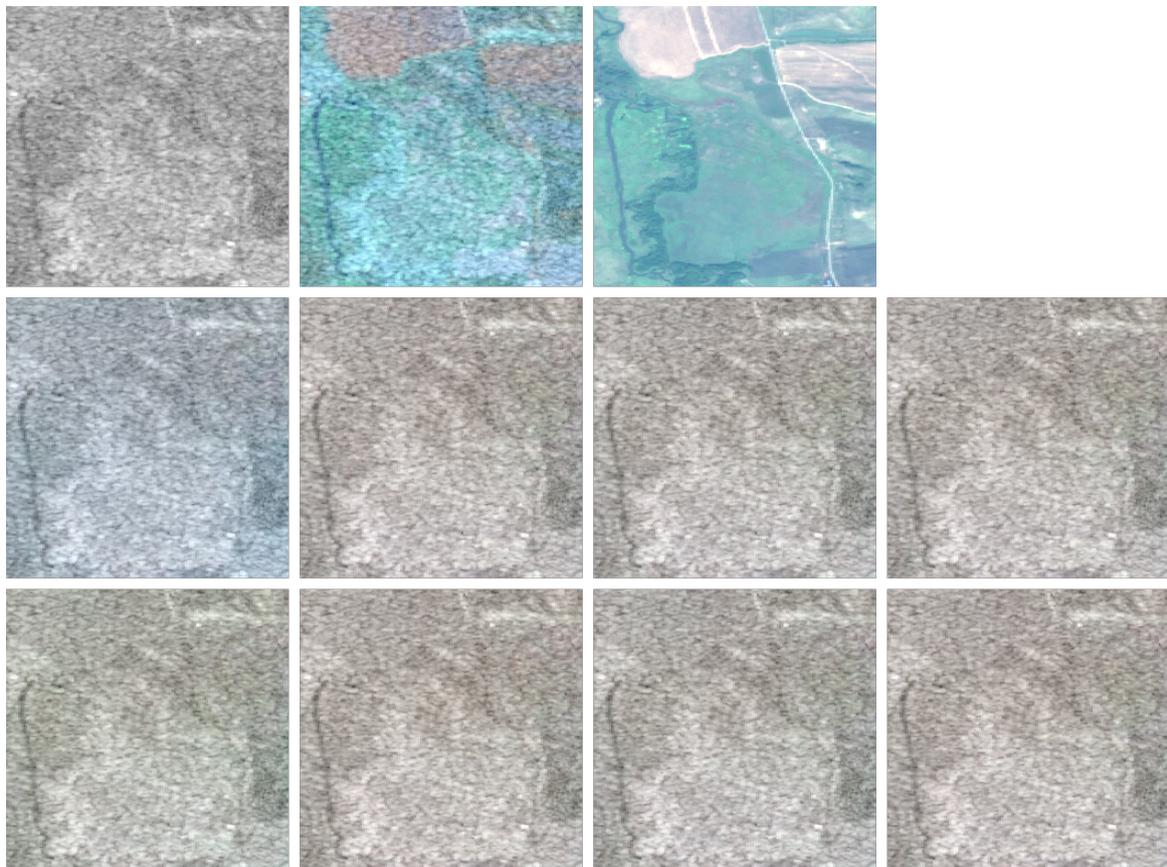
(c)



(d)



(e)



(f)

Figure 4. Some test-time examples for colorized Sentinel-1 SAR images. Top row in every subfigure: SAR image (left), target *Lab* image (center), corresponding optical image (right). The remaining 8 images in each subfigure show the top-8 colorization samples drawn from the learned color distribution.