https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-049

Deep Dimension Reduction for Spatial-Spectral Road Scene Classification

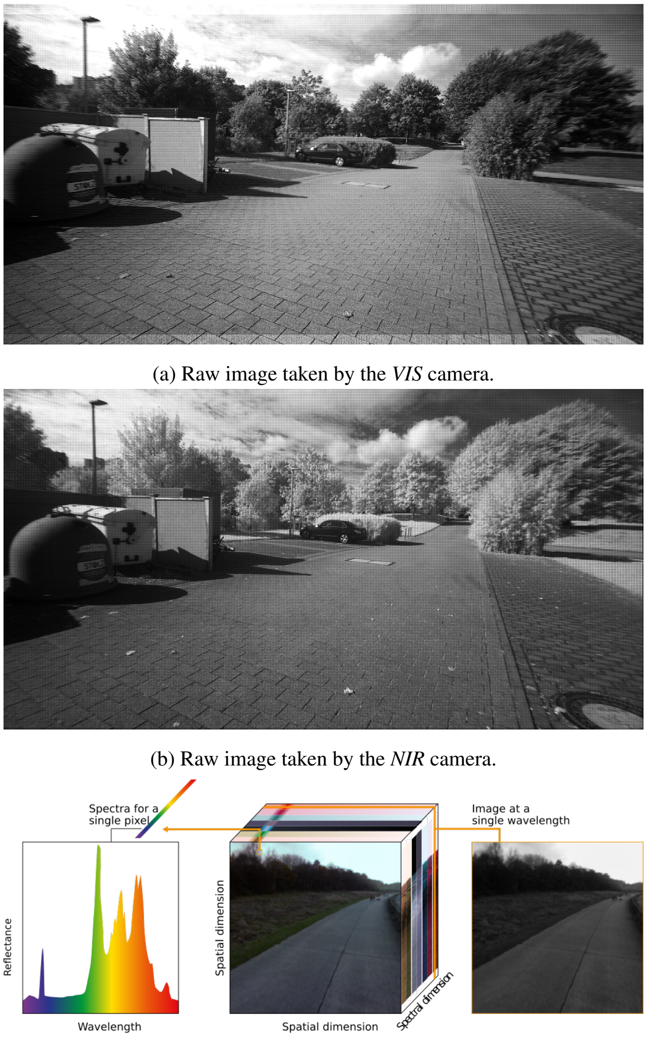
Christian Winkens, Florian Sattler, Dietrich Paulus, Active Vision Group, Institute for Computational Visualistics, University of Koblenz-Landau, Germany

Abstract

Semantic segmentation is an essential aspect of modern autonomous driving systems, since a precise understanding of the environment is crucial for navigation. We investigate the eligibility of novel snapshot hyperspectral cameras –which capture a whole spectrum in one shot– for road scene classification. Hyperspectral data brings an advantage, as it allows a better analysis of the material properties of objects in the scene. Unfortunately, most classifiers suffer from the Hughes effect when dealing with high-dimensional hyperspectral data. Therefore we propose a new framework of hyperspectral-based feature extraction and classification. The framework utilizes a deep autoencoder network with additional regularization terms which focus on the modeling of latent space rather than the reconstruction error to learn a new dimension-reduced representation. This new dimension-reduced spectral feature space allows the use of deep learning architectures already established on RGB datasets.

Introduction

Environmental perception and analysis are crucial for autonomous driving, especially in off-road scenarios. The correct semantic interpretation of a scene is a vital factor for successful autonomous navigation. In this case the application of hyperspectral sensors offers advantages as they provide a more detailed view of the composition and surface of materials, plants and ground materials than conventional cameras. These sensors capture spectral information over a specific part of the electromagnetic spectrum, turning the hyperspectral image (HSI) into a 3D datacube with spatial-spectral properties. Therefore, HSIs have been used in many applications, such as environmental monitoring, earth observation, object recognition, agriculture, etc. assuming that each scene is static. Because a big disadvantage of established sensors are the scan requirements for the construction of a hyperspectral cube (hypercube) of a scene like in displayed in figure 1. This leads to slow acquisition rates and motion artifacts when observing dynamic scenes such as driving scenarios. However, this disadvantage can be overcome with novel snapshot mosaic (SSM) sensors [Geelen et al., 2014], which can capture multiple spectra in one image and can therefore be used on unmanned land vehicles to provide hyperspectral classification of dynamic scenes.


(a) Raw image taken by the *VIS* camera.

(b) Raw image taken by the *NIR* camera.

(c) A schematic representation of a hypercube.

Figure 1: Examples of raw hyperspectral images taken with *VIS* and *NIR* camera with visible mosaic pattern. And a schematic representation of a hypercube used in this work.

To effectively use this data, supervised hyperspectral image classification is one of the most active research areas in hyperspectral data analysis. A special target function is calculated here to enable the learning of characteristics from annotated data. Recently, the great success of Deep Neural Networks (DNNs) in various computer vision tasks has been observed. Therefore, in some studies DNNs were trained using HSI data, where the desired features and the classifier are integrated into a uniform mapping function that can be learned together. This requires extensive, annotated data for training. However, annotating pixels in an HSI is costly and time consuming. Therefore, procedures must be developed that can be trained with limited annotated data. A common approach is to use unsupervised feature learning methods where features are extracted unattended and then integrated into a supervised classifier. Inspired by the idea of deep learning and the new SSM hyperspectral sensors, we propose a HSI spectral-spatial classification system based on dimension reduction and deep features. First we train an autoencoder (AE) with custom regularizations focusing on modeling the latent space rather the reconstruction error to learn a new dimension-reduced representation of hyperspectral data. The learned representation is then used as input data for HSI spatial-spectral classification using deep convolution neural networks (DCNNs). Our goal is to examine the use of hyperspectral data and deep learning for dynamic scene understanding in autonomous driving scenarios.

The remainder of this paper is organized as follows. Section introduces related works with an overview of common algorithms for feature extraction and spectral-spatial classification. Then our general sensor setup is presented in section . Our feature extraction and classification approach is described in detail in section . And in the section we present our experiments and results utilizing our new hand-labeled dataset. Finally a conclusion of our work is given in the last section.

Related Work

The standard procedure in the field of image-based scene segmentation is characterized by taking RGB images and trying to distinguish different classes, as Chetan et al. [Chetan et al., 2010] and others demonstrate. But in recent years hyperspectral imaging and classification has gained in importance. Hyperspectral data provide a more in-depth view of the structure and composition of objects and materials such as plants and soils than RGB data. Given hyperspectral data, the aim of classification is to assign a unique name to each spectral vector so that it is clearly defined by a certain class. Conventional hyperspectral classifiers use only spectral information, and the classification algorithms typically include methods like k-nearest neighbors, maximum probability, minimum distance, and logistic regression like mentioned by Foody et al. [Foody and Mathur, 2004]. Unfortunately, most of supervised classifiers monitored, suffer from the Hughes effect [Hughes, 1968], especially when dealing with high-dimensional hyperspectral data. There is an exponential growth of samples that are needed to maintain statistical confidence as dimensions grow. Furthermore there are other critical problems in the classification of hyperspectral data, such as the limited number of annotated data and large spatial variability of spectral signatures [Camps-Valls and Bruzzone, 2005]. To compensate for the high dimensionality and the limited amount of annotated data [Ambikapathi et al., 2013], some techniques were developed which focus on dimension reduction. Alternatively dimension transformation[Jimenez and Landgrebe, 1999, Harsanyi and Chang, 1994, Bruce et al., 2002] and band selection [Samadzadegan et al., 2012, Chang et al., 1999, Serpico and Bruzzone, 2001] are ways for dealing with high dimensionality. SVM classifiers have long been the most modern methods [Zhuo et al., 2008] for hyperspectral classification, as they have a low sensitivity to high dimensionality and do not suffer from the Hughes phenomenon [Gualtieri and Chettri, 2000]. In order to deal with the spatial variability of the spectral signature, some recent approaches try to take into account spatial information as proposed by Tarabalka et al. [Tarabalka et al., 2009] and Plaza et al. [Plaza et al., 2009], which has gained in importance in recent years. It has been shown that these methods allow significant performance improvements in classification. Plaza et al. [Plaza et al., 2009] presented a method based on the fusion of morphological information and original data, followed by an SVM and providing good classification results. Several spectral-spatial classifiers take into account the spatial smoothness based on the pixel-wise classification, e.g. the use of a segmentation map to regularize the pixel-wise classification result, as proposed by Huang et al. and Tarabalka [Huang and Zhang, 2009, Tarabalka, 2010]. Furthermore Li et al. [Li et al., 2013] proposed a new approach which uses spatial and spectral information with the help of loopy belief propagation and active learning. Recently, deep learning utilizing neural networks has proven to be promising in many areas, including classification or regression tasks, especially with the use of images [Krizhevsky et al., 2012, Hinton and Salakhutdinov, 2006, Zuo and Wang, 2014]. Neural networks learn to represent features through a multi-level training process and can learn structures by minimizing the mean square error of all samples from different classes. The hierarchical structure resembles the recognition process of the human brain, which is able to learn more abstract concepts than hardcoded feature representation methods. Furthermore Mou et al. [Mou et al., 2017] proposed a novel neural network with a new activation function and a modified gated recurrent unit for hyperspectral image classification, which analyze hyperspectral pixels as sequential data. The networks mentioned above are 1-D deep learning architectures that are equipped with fully connected layers. In comparison, a convolutional neural network (CNN) uses local connections to deal with spatial dependencies via sharing weights and can thus significantly reduce the number of parameters of the network compared to conventional 1-D fully connected neural networks. CNNs have already surpassed other methods in various areas, such as scene understanding [Long et al., 2015, Noh et al., 2015]. Furthermore a few supervised CNNs for hyperspectral-spatial classification have been proposed. A regularized 3-D CNN-based feature extraction model to extract efficient spectral-spatial features was introduced by Chen et al. [Chen et al., 2016]. Ghamisi et al. [Ghamisi et al., 2016] combined a CNN with a fractional order darwinian particle swarm optimization algorithm to iteratively select the most informative bands for training in hyperspectral data. Despite the great success of the supervised CNNs, there is a need for a good supply of labeled training samples, which unfortunately are difficult to obtain. So supervised CNNs usually suffer from a small number of training samples or unbalanced data sets. For this reason, the unsupervised learning of spectral-spatial features, which has quick access to any amount of unlabeled data, is conceptually of great interest. The main purpose of unsupervised feature learning is to extract essential features from unlabeled data, detect and remove input redundancies and obtain only key aspects of the data in robust and discriminatory representations.

Lin et al. [Lin et al., 2013] and Chen et al. [Chen et al., 2014] proposed stacked autoencoders to extract hierarchical features from the spectral range of hyperspectral images for classification. Furthermore Zhao et al. [Chen et al., 2015] presented a multi-scale, stacked autoencoder to learn an effective feature representation from unlabeled data combined with a linear SVM for hyperspectral data classification. Later this scheme was improved by

Tao et al. [Tao et al., 2015] who proposed an autoencoder which learns an overcomplete sparse feature representation, which tends to be more effective and discriminative for classification. Recently Wang et al. [Wang et al., 2018] presented a novel low-rank representation based HSI classification framework where the unsupervised learning scheme and the robust classification are modelled separately.

Sensors

The sensors used in this work utilize a specific filter mosaic structure, which has a per-pixel design developed by IMEC [Geelen et al., 2014]. The filters are arranged in a rectangular mosaic pattern of n rows and m columns, which is repeated w times over the width and h times over the height of the sensor. We used two different camera models from Ximea, the MQ022HG-IM-SM4X4-VIS (*VIS*) which captures the visible spectrum 470–630 nm and the MQ022HG-IM-SM5X5-NIR (*NIR*) which is designed for the near-infrared range 600-975 nm. The *VIS* camera has a 4×4 mosaic pattern and the *NIR* 5×5 which results in a spatial resolution of approx. 512×272 pixels (4×4) and 409×217 pixels (5×5). The cameras provide images in a lossless format with 8 bits per sample. Therefore the raw data captured by the camera needs a special preprocessing to construct a hypercube with spectral reflectances from the raw data like seen in figure 1. Preprocessing consists of cropping the raw-image to the valid sensor area, removing the vignette and converting to a three dimensional image, which we call a hypercube, like described in [Winkens et al., 2017].

Semantic Scene Analysis

Autoencoder for dimension reduction

An autoencoder [Hinton and Salakhutdinov, 2006, Vincent et al., 2010] is a symmetrical neural network that allows the properties of a dataset to be learned unsupervised. It takes an input $x \in \mathbb{R}^D$ and maps it to a latent representation $h \in \mathbb{R}^M$ using a nonlinear mapping $h = f(\omega x + \beta)$ where β is a bias vector and ω defines a weight matrix which needs to be trained. Furthermore f defines a nonlinear activation function such as a sigmoid function. A reverse mapping $y = f(\omega' h + \beta')$ is used to reconstruct the input data x from the latent representation h with $\omega' = \omega^T$. If $M < D$, the autoencoder is called *undercomplete* and learns a low-dimensional compressed data depiction, representing the most salient features of the data distribution. The learning process minimizes the reconstruction error

$$\mathrm{L} = \frac{1}{n}\sum_{i=1}^{n}(x-y)^2 \tag{1}$$

and as a side-effect a latent space is constructed. If the autoencoder is linear and the loss function is defined as the mean squared error like in equation 1, the autoencoder spans a subspace comparable to a principal component analysis (PCA). Learning an undercomplete representation forces the autoencoder to capture the most salient features of the training data.

Autoencoders are often trained with a single layer encoder and a single layer decoder only [Goodfellow et al., 2016]. But the construction of deep autoencoder networks with several hidden layers can offer many advantages such as higher robustness to noise, capturing of non linear relationships and generally a superior function approximation. The universal approximator theorem [Csáji, 2001] guarantees that a neural feedforward network with at least one hidden layer can represent an approximation of any function with an arbitrary degree of accuracy, assuming that it has enough hidden units. So-called regularized autoencoders use a loss function that stimulates the model to learn different properties other than the ability to copy the input into its output. These properties, among others, include denoising, missing inputs or sparsity of the representation.

In contrast to common procedures where the primary focus is on minimizing the reconstruction error, our focus is on constructing a latent space that enables a meaningful compressed representation of the hyperspectral data. We have taken various measures to achieve this, which we explain in the following. The latent space is explicitly conditioned by introducing special-regularizations.

To shape this space the batch size needs to be considered as it solely determines the shape of the latent data distribution. Autoencoders are generally trained using mini batches and stochastic gradient descent (SGD). The mini batch size depends on the specific task. Smaller sizes lead to rapid changes while larger sizes consider more data and change the network slower. Here the batch size must be large enough to make a statistical statement but has to be small enough to find a suitable solution.

We use a structured loss S to shape the latent distribution which is defined as

$$S = g(\bar{x} - \mu^*) + g(\sigma_X - \sigma^*) \tag{2}$$

where g computes the sum of squared elements across the first dimension of a tensor and $\bar{x}, \sigma_X$ define the mean and standard deviation of x. The term σ^* denotes the desired standard deviation and similar to this μ^* denotes the desired mean. This formulation of the loss ensures that the statistical properties are present for each latent dimension and are not just valid for the overall space. Furthermore, we introduce a weight decay $W = \sum_i^n |w_i|$ to prevent the network from overfitting. This term limits the growth of network weights and simultaneously reduces the number of free weight parameters. It leads to a well defined model and is an example for a common sparsity constraint. The overall cost of the training task is defined as $C = \alpha_0 \cdot L + \alpha_1 \cdot S + \alpha_2 \cdot W$ where $\alpha_0 \gg \alpha_1 \gg \alpha_2$. This forces the training process to first minimize the reconstruction error and as soon as this problem is solved with sufficient accuracy, the distribution of the latent representation is optimized. Overall, the growth of the weights is limited and it can even be reduced over time if no other improvement to the model is possible otherwise.

Utilizing these concepts we construct our architecture with three encoding operations and three decoding operations which share their weights. As depicted in figure 2 each layer has a hyperbolic tangent activation. This keeps the outputs of every layer, even the network outputs in a defined range. As the last layer has also a sigmoid activation, target values of 1 or -1 would lead to increasing weights. To keep the weights in the last layer from exploding, we scale the input data to be in range from -0.5 to 0.5. With respect to this activation the latent space is conditioned to have $\mu^* = 0$ and $\sigma^* = 0.1$.

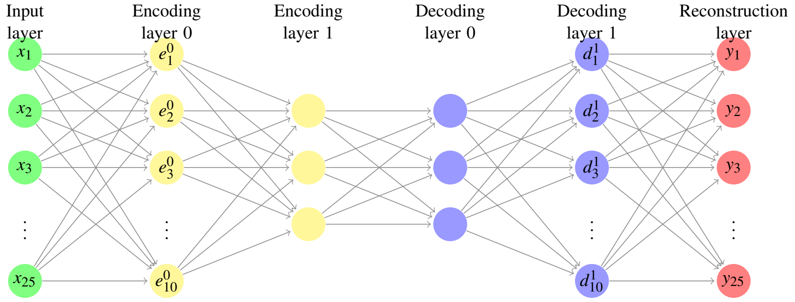
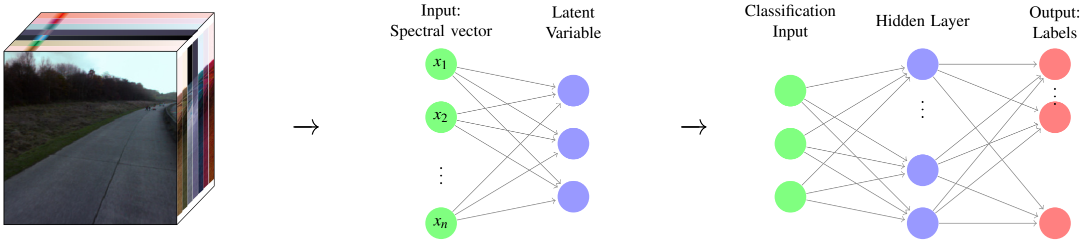


Figure 2: Instance of our autoencoder for hyperspectral NIR data with 25 channels. The latent space has a dimension of 3.



Input: hypercube

Our autoencoder network.

CNN architecture as semantic classifier.

Figure 3: Scheme of our classification framework, which has two parts: A spectral hyperpixel vector as input for the autoencoder. The latent space of the autoencoder is reassembled to an image with 3 channels and serves as input for the CNN with an output layer for semantic segmentation.

Name	IoU Cityscapes	Hyperspectral	Year	Ref
Encoder-Decoder (SegNet)	57.0	yes	2015	[Badrinarayanan et al., 2015]
DenseNet	NA	yes	2016	[Jégou et al., 2016]
PSPNet	81.2	no	2016	[Zhao et al., 2016]
RefineNet	73.6	no	2016	[Lin et al., 2016]
FRRN	71.8	yes	2016	[Pohlen et al., 2017]
MobileUNet	NA	yes	2017	[Howard et al., 2017]
DeepLabv3	82.1	no	2017	[Chen et al., 2017]
GCN	80.5	no	2017	[Peng et al., 2017]
DenseASPP	80.6	no	2018	[Yang et al., 2018]
BiSeNet	68.4	no	2018	[Yang et al., 2018]

Table 1: Overview of tested and trained network architectures.

Neuronal Nets for Semantic Segmentation

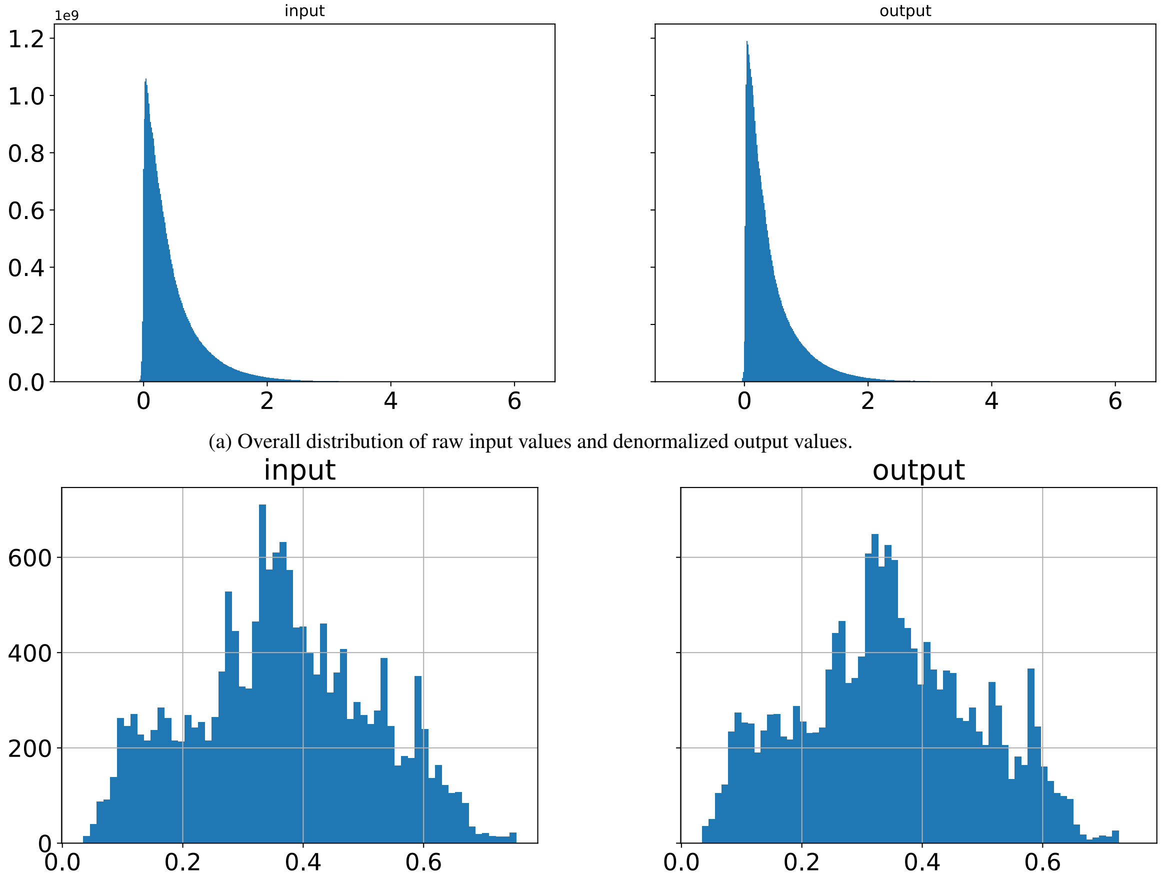
The most successful semantic segmentation approaches in recent years have been relying on convolutional neural networks (CNNs). Pixel-wise classification was performed using CNN features from various scales, followed by aggregation of noisy pixel predictions across superpixel regions. Long et al. [Long et al., 2015] introduced fully convolutional networks (FCNs) for semantic segmentation which opened up a new spectrum of analysis through end-to-end training. We want to investigate the effectiveness of our autoencoder architecture in combination with different neuronal networks. Furthermore, we want to make a statement about which data provides the best results. For this purpose we have created a dataset in which the data from *VIS* and *NIR* camera was recorded synchronously. We then extracted and annotated the synchronized hypercubes from both cameras. This allows us to investigate which data (*VIS*,RGB,*NIR*) representation is most effective for scene segmentation. In table 1 an overview of network architectures we have tested with our autoencoder is given. Unfortunately not all architectures are able to process data with more than three channels without extensive changes to their structure. This is also due to the fact that they use a pre-trained ResNet in the frontend, which relies on RGB data and is not compatible with our high dimensional data. The networks which can be trained high dimensional data are marked accordingly. With regard to table 1 we trained four networks with hyperspectral (*VIS*,*NIR*) data and the others were trained using RGB or compressed (autoencoder) data.

Experiments

As far as we know, there is no publicly available set of hyperspectral data recorded by the MQ022HG-IM-SM4X4-VIS camera and MQ022HG-IM-SM5X5-NIR camera that use snapshot mosaic technology (SSM) to capture hyperspectral data. So we had to create a new dataset ourselves, which includes several hundreds annotated hypercubes and is publicly available. We equipped a standard car with the cameras manufactured by Ximea and collected several hours of data where we drove through suburbs and rural areas, from which we selected a subset for labeling hyperspectral data. Therefore we have published a freely available synchronized and calibrated autonomous driving dataset, which covers different scenarios. To the best of our knowledge it is the first dataset providing snapshot mosaic hyperspectral data from the visible to the near infrared range. We offer semantic labels for synchronized *VIS* and *NIR* data to investigate the use of hyperspectral data for semantic scene understanding especially in autonomous driving scenarios.

Our Autoencoder Network

To train our autoencoder network we have extracted more than $1,000,000,000$ hyperpixels from our datasets. The autoencoder itself receives a single hyperpixel vector with 16 (VIS) or 25 (NIR) channels as input. To specify a fixed value range, the hyperpixel vectors are normalized accordingly so that every channel has a value range between -0.5 and 0.5. This means that the mean value is about 0, so the latent space is already given a certain structure. This normalization per channel is very important, because normalization over the whole vector could lead to bad conditioning of the individual channels. The training was carried



(a) Overall distribution of raw input values and denormalized output values.

(b) Histogram with the distribution of the mean values overall.

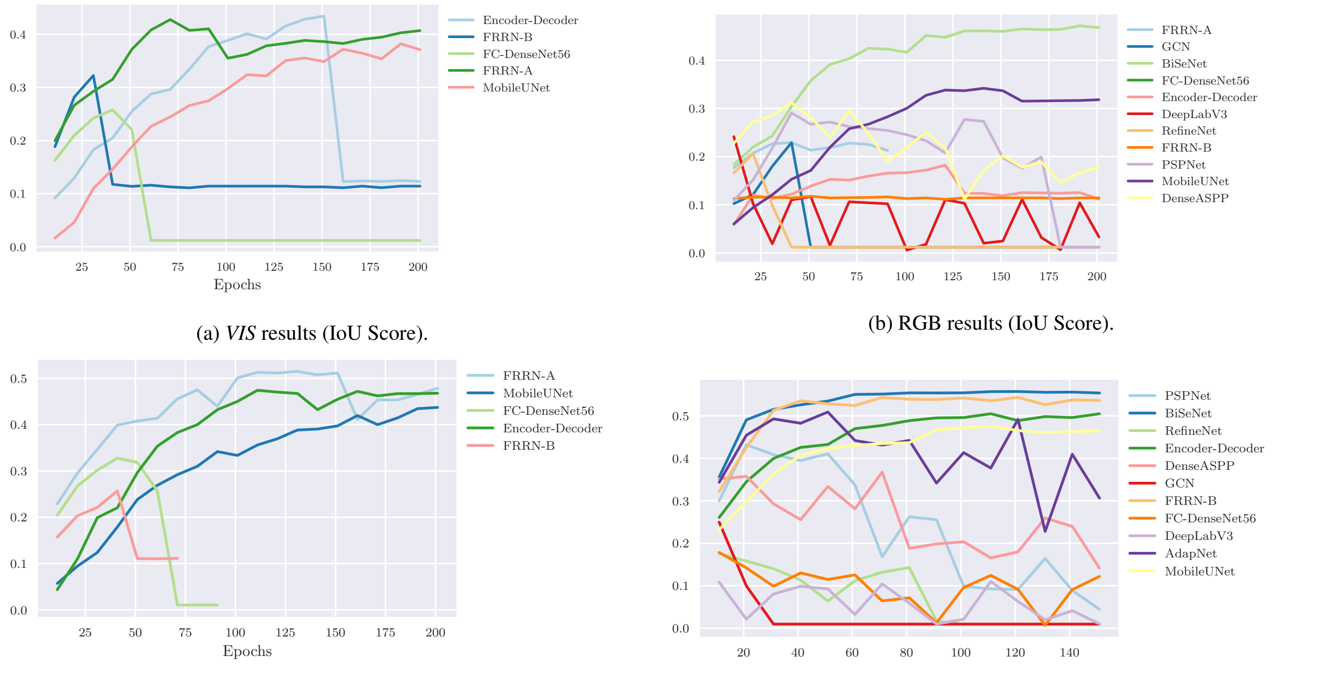
Figure 4: Results of our autoencoder training evaluation based on $1,000,000,000$ hyperspectral vectors from the *NIR* camera.

out with a batch size of 100000 and a learning rate of 0.001. Our training results for the autoencoder based on *NIR* data are displayed in figure 4. The overall input and output value distribution shows that the autoencoder was able to compress and reconstruct the input data very well. Here only small deviations in the distribution are to be recognized. In figure 4b a histogram with 64 bins displays the distribution of the mean values from input and output data. This also indicates that our autoencoder network makes only few errors reconstructing the given data It can therefore be concluded that the autoencoder has built up a very efficient three-dimensional feature space of the 25-dimensional input data. This clearly indicates the precision of our autoencoder.

Deep Neural Network

Since we want to classify hyperpsectral data, we can unfortunately not use pre-trained nets. Therefore we have to train the network weights from scratch with only limited data. As described above, our pre-trained autoencoder is connected to different networks with fixed autoencoder weights, so that only the neural network weights are optimized during the training. Whereas the autoencoder learned a three-dimensional feature space from the input data in the first stage. We report the results of our combined network trained on nearly 200 annotated hypercubes with semantic annotations including 10 classes on a resolution of 256×512 pixels. Since our hyperspectral data set cannot keep up with data sets like Cityscapes in size, we use data augmentation to increase the data set artificially. The network was trained for 200 iterations with a batch size of four by minimizing a cross-entropy loss utilizing RMSProp optimizer.

In figure 5a the results of the training with the 16 channel *VIS* data are displayed. As mentioned above, only a part of the networks can be trained directly with the hypercubes. In the end three networks could achieve reasonable results. DenseNet and SegNet did not work well and after a short time they did not train anything meaningful anymore. Looking at the results, it turned out that everything was simply classified as street. In order to make a targeted comparison possible, we generated RGB data from the 16 *VIS* channels, the results are displayed in figure 5b. Most of the networks show the same behavior as the previous experiment and classify almost everything as street after a short time. In contrast, BiseNet and MobileUNet achieved good results. In figure 5c the results of the experiments with raw *NIR* data are displayed. Overall, the results of the networks FRRN-A, MobileUNet and Encoder-Decoder are better than those trained with *VIS* and the RGB data. As before, DenseNet can do little with hyperspectral data and has not trained anything useful. Finally, we combined the available architectures with our pre-trained autoencoder architecture. Since our autoencoder has a three-dimensional latent space, we can train all selected networks. The results using compressed *NIR* data are better than with raw *NIR* data. It is also noticeable that the networks that showed superb results on other data representations are also ahead. The best architecture using compressed



(a) *VIS* results (IoU Score).

(b) RGB results (IoU Score).

(c) NIR Results (IoU Score)

(d) AE NIR Results (IoU Score)

Figure 5: Results of our evaluation on our dataset.

NIR data is the BiseNet which already showed good results on the RGB data. It clearly shows that the combination of Autoencoder and *NIR* data in combination with BiSeNet achieves the best results. This is followed by the unprocessed *NIR* data. The *VIS* data is the worst performer here. RGB data show significantly better results with BiseNet than the unprocessed 16 *VIS* channels representation.

Conclusion

We proposed a spatial-spectral feature learning framework for HSI classification which combines unsupervised and supervised deep learning methods. To better characterize each hyperpixel in the spectral space, we proposed the unsupervised learning of a deep autoencoder with additional regularization terms which focus on the modeling of latent space rather the reconstruction error to learn a new dimension-reduced representation. This new feature space allows the use of deep learning methods and networks already showing impressive results on RGB data. Therefore we adopted deep learning methods by testing differen convolutional neural networks for spatial and spectral road scene classification. Experiments were carried out on our novel hyperspectral ground truth dataset which is freely available. The results and reconstruction error of the trained autoencoder show promising robustness and transferability of the learned features. The combination of our autoencoder network and established deep learning classifiers leads to an accurate pixel-level classification performance as our experiments indicate. Our future work involves incorporating different deep learning architectures into our framework to further improve the classification accuracy. Additionally we intend to fuse 3D laser data with the hyperspectral data in a next step.

References

[Ambikapathi et al., 2013] Ambikapathi, A., Chan, T.-H., Lin, C.-H., and Chi, C.-Y. (2013). Convex geometry based outlier-insensitive estimation of number of endmembers in hyperspectral images. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2013 5th Workshop on*, pages 1–4. IEEE.

[Badrinarayanan et al., 2015] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561.

[Bruce et al., 2002] Bruce, L. M., Koger, C. H., and Li, J. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on geoscience and remote sensing*, 40(10):2331–2338.

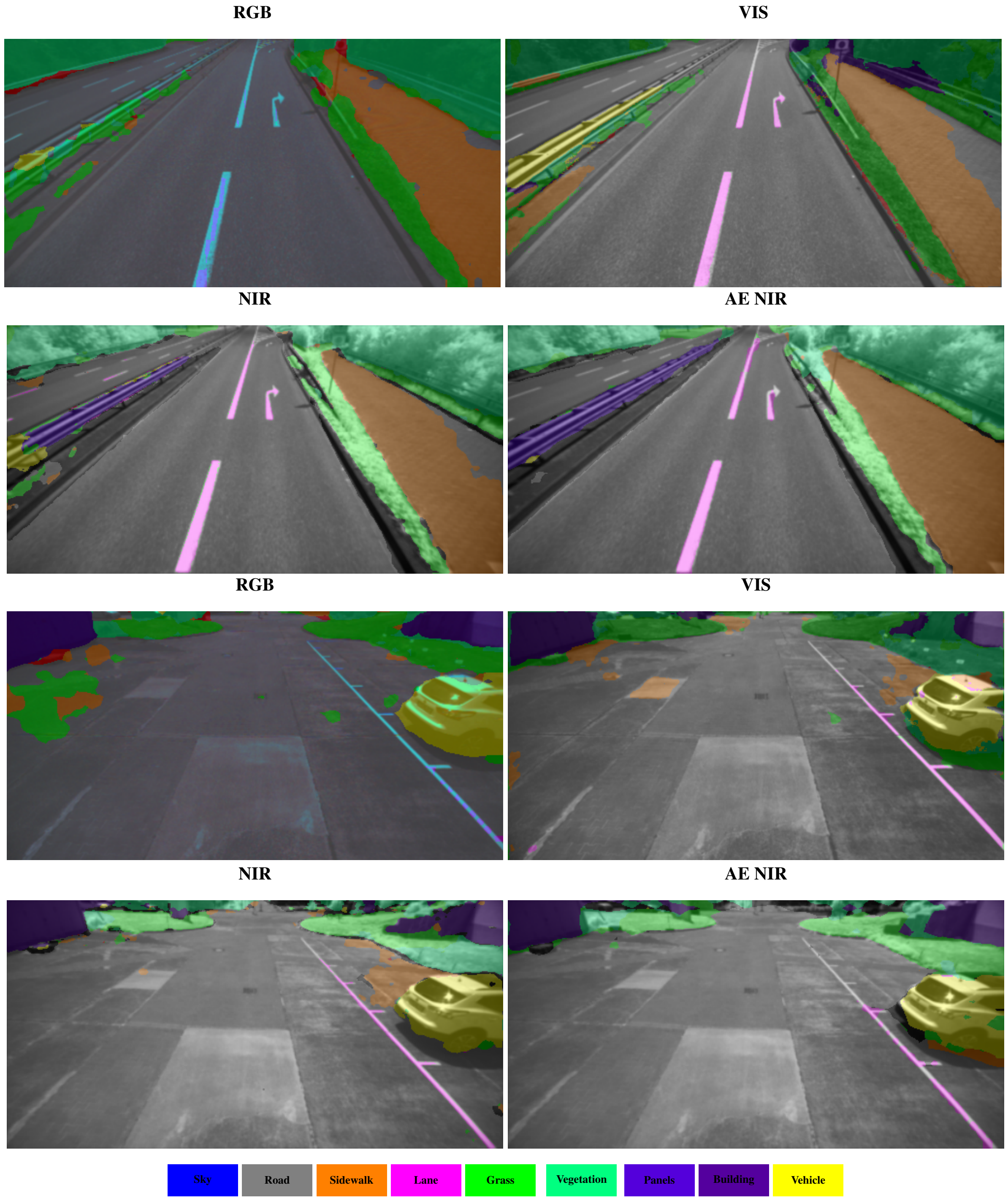
[Camps-Valls and Bruzzone, 2005] Camps-Valls, G. and Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362.

[Chang et al., 1999] Chang, C.-I., Du, Q., Sun, T.-L., and Althouse, M. L. (1999). A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE transactions on geoscience and remote sensing*, 37(6):2631–2641.

[Chen et al., 2017] Chen, L., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587.

[Chen et al., 2016] Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251.

[Chen et al., 2014] Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107.



[Chen et al., 2015] Chen, Y., Zhao, X., and Jia, X. (2015). Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392.

[Chetan et al., 2010] Chetan, J., Krishna, M., and Jawahar, C. (2010). Fast and spatially-smooth terrain classification using monocular camera. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 4060–4063. IEEE.

[Csáji, 2001] Csáji, B. C. (2001). Approximation with artificial neural networks. *Faculty of Sciences, Etvs Lornd University, Hungary*, 24:48.

[Foody and Mathur, 2004] Foody, G. M. and Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(6):1335–1343.

[Geelen et al., 2014] Geelen, B., Tack, N., and Lambrechts, A. (2014). A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic. In *Spie Moems-Mems*, pages 89740L–89740L. International Society for Optics and Photonics.

[Ghamisi et al., 2016] Ghamisi, P., Chen, Y., and Zhu, X. X. (2016). A self-improving convolution neural network for the classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 13(10):1537–1541.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[Gualtieri and Chettri, 2000] Gualtieri, J. and Chettri, S. (2000). Support vector machines for classification of hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, volume 2, pages 813–815. IEEE.

[Harsanyi and Chang, 1994] Harsanyi, J. C. and Chang, C.-I. (1994). Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Transactions on geoscience and remote sensing*, 32(4):779–785.

[Hinton and Salakhutdinov, 2006] Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507.

[Howard et al., 2017] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861.

[Huang and Zhang, 2009] Huang, X. and Zhang, L. (2009). A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy. *International Journal of Remote Sensing*, 30(12):3205–3221.

[Hughes, 1968] Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63.

[Jégou et al., 2016] Jégou, S., Drozdzal, M., Vázquez, D., Romero, A., and Bengio, Y. (2016). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *CoRR*, abs/1611.09326.

[Jimenez and Landgrebe, 1999] Jimenez, L. O. and Landgrebe, D. A. (1999). Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

[Li et al., 2013] Li, J., Bioucas-Dias, J. M., and Plaza, A. (2013). Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):844–856.

[Lin et al., 2016] Lin, G., Milan, A., Shen, C., and Reid, I. D. (2016). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. *CoRR*, abs/1611.06612.

[Lin et al., 2013] Lin, Z., Chen, Y., Zhao, X., and Wang, G. (2013). Spectral-spatial classification of hyperspectral image using autoencoders. In *Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on*, pages 1–5. IEEE.

[Long et al., 2015] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[Mou et al., 2017] Mou, L., Ghamisi, P., and Zhu, X. X. (2017). Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3639–3655.

[Noh et al., 2015] Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1520–1528.

[Peng et al., 2017] Peng, C., Zhang, X., Yu, G., Luo, G., and Sun, J. (2017). Large kernel matters - improve semantic segmentation by global convolutional network. *CoRR*, abs/1703.02719.

[Plaza et al., 2009] Plaza, A., Plaza, J., and Martin, G. (2009). Incorporation of spatial constraints into spectral mixture analysis of remotely sensed hyperspectral data. In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*, pages 1–6. IEEE.

[Pohlen et al., 2017] Pohlen, T., Hermans, A., Mathias, M., and Leibe, B. (2017). Full-resolution residual networks for semantic segmentation in street scenes. *arXiv preprint*.

[Samadzadegan et al., 2012] Samadzadegan, F., Hasani, H., and Schenk, T. (2012). Simultaneous feature selection and svm parameter determination in classification of hyperspectral imagery using ant colony optimization. *Canadian Journal of Remote Sensing*, 38(2):139–156.

[Serpico and Bruzzone, 2001] Serpico, S. B. and Bruzzone, L. (2001). A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1360–1367.

[Tao et al., 2015] Tao, C., Pan, H., Li, Y., and Zou, Z. (2015). Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience and remote sensing letters*, 12(12):2438–2442.

[Tarabalka, 2010] Tarabalka, Y. (2010). *Classification of hyperspectral data using spectral-spatial approaches*. PhD thesis, Institut National Polytechnique de Grenoble-INPG.

[Tarabalka et al., 2009] Tarabalka, Y., Benediktsson, J. A., and Chanussot, J. (2009). Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2973–2987.

[Vincent et al., 2010] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408.

[Wang et al., 2018] Wang, C., Zhang, L., Wei, W., and Zhang, Y. (2018). When low rank representation based hyperspectral imagery classification meets segmented stacked denoising auto-encoder based spatial-spectral feature. *Remote Sensing*, 10(2):284.

[Winkens et al., 2017] Winkens, C., Sattler, F., and Paulus, D. (2017). Hyperspectral terrain classification for ground vehicles. In *12th International Conference on Computer Vision Theory and Applications (VISAPP)*.

[Yang et al., 2018] Yang, M., Yu, K., Zhang, C., Li, Z., and Yang, K. (2018). Denseaspp for semantic segmentation in street scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition*.

[Zhao et al., 2016] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2016). Pyramid scene parsing network. *CoRR*, abs/1612.01105.

[Zhuo et al., 2008] Zhuo, L., Zheng, J., Li, X., Wang, F., Ai, B., and Qian, J. (2008). A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine. In *Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images*, volume 7147, page 71471J. International Society for Optics and Photonics.

[Zuo and Wang, 2014] Zuo, Z. and Wang, G. (2014). Learning discriminative hierarchical features for object recognition. *IEEE Signal Processing Letters*, 21(9):1159–1163.

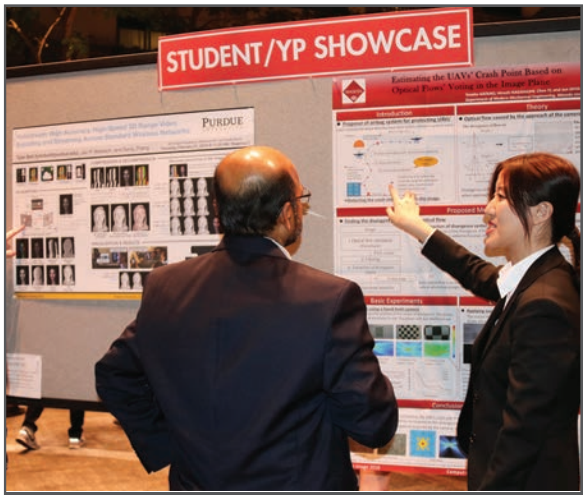
JOIN US AT THE NEXT EI!

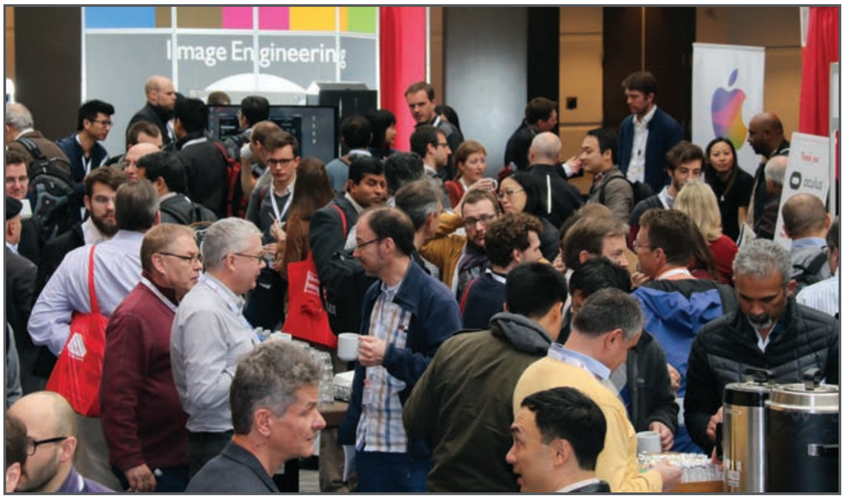
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





- AR is a fundamental and transformative interface and platform.
- 20+ years from my initial working demo to today's AR displays
- Optimistic that it won't take another 20 to reach ubiquitous consumer AR!

• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org