Utilizing Continuous Kernels for Processing Irregularly and Inconsistently Sampled Data With Position-Dependent Features

Birk Martin Magnussen
Opsolution GmbH
Ziegelstraße 17
34121 Kassel, Germany
e-mail: birk.magnussen@opsolution.de

Claudius Stern
FOM Hochschule für Oekonomie & Management
Hochschulzentrum Kassel
Garde-du-Corps-Straße 7
34117 Kassel, Germany
e-mail: Claudius.Stern@fom.de

Bernhard Sick
Intelligent Embedded Systems
Universität Kassel
Wilhelmshöher Allee 73
34121 Kassel, Germany
e-mail: bsick@uni-kassel.de

***Abstract*—Continuous Kernels have been a recent development in convolutional neural networks. Such kernels are used to process data sampled at different resolutions as well as irregularly and inconsistently sampled data. Convolutional neural networks have the property of translational invariance (e.g., features are detected regardless of their position in the measurement domain), which is unsuitable for certain types of data, where the position of detected features is relevant. However, the capabilities of continuous kernels to process irregularly sampled data are still desired. This article introduces a novel method utilizing continuous kernels for detecting global features at absolute positions in the data domain. Through a use case in processing multiple spatially resolved reflection spectroscopy data, which is sampled irregularly and inconsistently, we show that the proposed method is capable of processing such data natively without additional preprocessing as is needed using comparable methods. In addition, we show that the proposed method is able to achieve a higher prediction accuracy than a comparable network on a dataset with position-dependent features. Furthermore, a higher robustness to missing data compared to a benchmark network using data interpolation is observed, which allows the network to adapt to sensors with individual failed components without the need for retraining.**

***Index Terms*—machine learning; neural nets; continuous kernel; irregularly sampled data; reflection spectroscopy**

I. Introduction

Common machine learning methods assume that data is sampled consistently. That is, each instance of sampled data has the same shape and each data point always represents the same value. However, in real-world applications, data might often be sampled inconsistently due to factors like production inaccuracies of sensors measuring the data. In some cases, certain data points may be missing from a measurement as well. To facilitate this type of data, imputation of missing data points can be used to reconstruct the data [1]. Some methods also employ neural networks to reconstruct or correct data [2]. For certain types of network architectures, such as convolutional neural networks [3], continuous kernels have been utilized to circumvent these assumptions and to handle irregularly and inconsistently sampled data natively instead of needing special preprocessing [4] [5] [6]. Convolutional neural networks have the property of translational invariance, which is ill-suited to data expected to exhibit features at consistent absolute positions within the data sampling domain. Common examples of this type of data include spectral data, both optical and acoustic, where the relevant features may be intensity peaks at specific, consistent wavelengths rather than a wavelength-invariant feature of the intensity curve's shape.

To process data with position-dependent features natively when sampled irregularly and inconsistently, a novel method is proposed in this paper. It has been shown that neural networks, such as sinusoidal representation networks (SIRENs) [7] can be used as functions parametrized through their learnable parameters, making them suitable for use as continuous kernels. These kernels are shown to be capable of modeling global, long-term dependencies [4]. By utilizing such continuous kernels outside of the context of convolutional neural networks, our method is capable of natively processing irregularly and inconsistently sampled data with position-dependent features.

The rest of the article is organized as follows. Section 2 details the current state-of-the-art regarding continuous kernels. Section 3 discusses the definition of a continuous kernel. Section 4 introduces the new methodology utilizing continuous kernels. Section 5 shows the efficacy of the proposed method for processing spectroscopy data. Section 6 discusses potential future research into using continuous kernels for explainable AI. The conclusion closes the article.

II. Related Work

Previous work on continuous kernels focuses primarily on applications in convolutional neural networks (CNNs) to handle irregularly sampled data. In [4], continuous kernels are utilized to process various types of sequential data, including irregularly sampled sequences. The article also performs an in-depth analysis of different types of continuous kernel parametrizations. [5] uses continuous kernels for convolutional neural networks to process non-grid bound data, such as representations of atoms in chemistry. In [6], continuous kernels are used to perform three-dimensional convolution on point clouds. However, as these articles all discuss the usage of continuous kernels for convolution in typical CNN architectures which all feature certain degrees of translational invariance, the methods discussed are not well suited for use with data containing position-dependent features.

Utilizing neural networks to represent a continuous function parametrized through the learnable parameters of the network, called implicit neural representation, has been previously analyzed by [8] [9] to model signed distance functions which are required for shape representation of 3D geometry. In [7], a network architecture called SIREN is proposed as an implicit neural representation for generic data, including audio, images, and signed distance functions.

III. Definition of Continuous Kernels

At its core, a continuous kernel is a function that assigns a weight to a data point at any given position [5]. Unlike in previous applications in CNNs, however, the position supplied to the kernel in our methods is an absolute position in the domain rather than a relative position to a convolution point.

To be able to represent continuous variants of typical weight kernels, the kernel function needs to be parametrizable with learnable parameters in such a way that the kernel function can ideally approximate any arbitrary function. It has been shown, that multi-layer perceptrons using sine nonlinearities, such as SIREN networks [7] can be used for such purposes.

Formalities

Small letters denote scalars, and small bold letters denote vectors. Capital letter variants of the former denote a set of the respective type. Subscripts on values indicate an index of the value within a containing set, superscripts indicate an index to the element of a vector.

Definition

Let $\mathbf{p}_i \in P \subset \mathbb{R}^n$ be the position of the value $d_i \in D \subset \mathbb{R}$ of the i-th data point of the set of data points D in an n-dimensional domain. A continuous kernel in the proposed architecture is now defined as a function

$$\psi : \mathbb{R}^n \mapsto \mathbb{R} \tag{1}$$

assigning a weight value to any position $\mathbf{p}_i$ in the domain. As shown in [4], such functions can be modeled and parameterized using implicit neural representations, such as Multi-Layer Perceptrons (MLP) using sine nonlinearities like SIREN [7]. In the proposed method, such an MLP serves as the function ψ. The MLP has n input neurons to input the absolute position $\mathbf{p}_i \in \mathbb{R}$ of a data point in the domain and one output neuron representing the assigned weight for the data point. The remaining model parameters of the kernel are the number and size of hidden layers in the MLP which can be adjusted to the problem to be learned. The MLP serving as the weight function ψ of a continuous kernel is not trained separately, but rather as part of the final network that the continuous kernel is used in.

IV. Application of Continuous Kernels in MLPs

Figure 1 shows the general structure of the proposed architecture. I shows the set of input data points to the model, each representing a value at a specific position within the measurement domain. In the proposed method, the first layer of the architecture, called the continuous feature layer, contains multiple independent continuous kernels (see II in Figure 1). For each of the independent kernels, the input consisting of an arbitrary number of data points is weighed using the kernel. Additionally, the input data might be sampled unevenly. To compensate for an uneven distribution of samples the local density of the sampled data points in the measurement domain is calculated (omitted in Figure 1). In the proposed method, kernel density estimation where the kernel size is a learnable parameter was used, but other methods for point density estimation can be used as well. Each data point is weighted by the inverse local density of data points at its position as proposed in [6]. The data points weighted by both the kernel and the inverse point density are shown in III in Figure 1 and are formally expressed in Equation 2. For each kernel, the weighted data points are reduced to a single value as defined in Equations 3 and 4 and as shown in IV of Figure 1. In the proposed method, a sum is used as the reduction operation, but other reductions, such as calculating the mean of the values are also considerable. Combining the reduced value of each

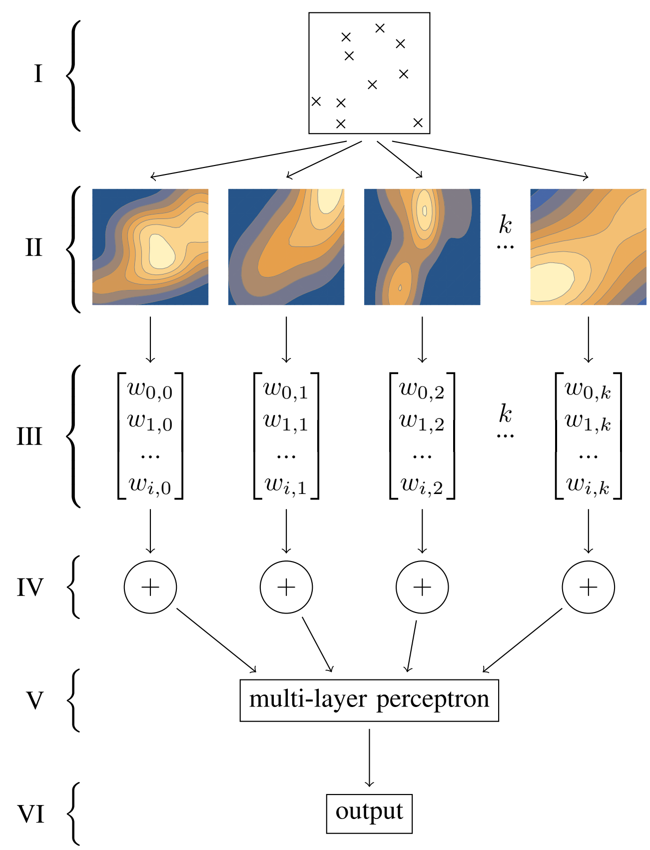

Fig. 1. Overview of a continuous feature network.

kernel into a vector results in an output feature vector of a fixed size depending on the number of independent kernels in the continuous feature layer. Since the continuous feature layer has reduced the input of arbitrary size to a latent vector of a fixed and predetermined size, the continuous feature layer can be followed with a typical neural network architecture, such as a multi-layer feed-forward network (see V in Figure 1). The output of this MLP then serves as the output of the entire network as depicted in VI of Figure 1. We call the proposed combination of a continuous feature layer followed by a multi-layer feed-forward network a continuous feature network. The continuous feature layer as described has three main model parameters: The number of kernels and the two parameters defining the shape of the kernels, being the number and the size of its hidden layers.

Formal Definition

Let the set Ψ be the set of multiple, independent continuous kernels ψ_k used in the continuous feature layer. In this set, each ψ_k represents one feature possibly present in the sampled data. Let $d_i \in D$ be the i-th data point in the input data with the position $\mathbf{p}_i \in P$ in the measurement domain. Let $\rho(\mathbf{p}_i)$ denote the local density of sampled data at position $\mathbf{p}_i$ in the domain. Then we define the weighted data points for each kernel as follows:

$$w_{i,k} := d_i \cdot \psi_k(\mathbf{p}_i) \cdot \frac{1}{\rho(\mathbf{p}_i)} \tag{2}$$

The components of the resulting fixed-size latent feature vector $\mathbf{v}$ are defined as follows:

$$\mathbf{v}^k(D,P) := \sum_i (w_{i,k}) \tag{3}$$

$$= \sum_i \left(d_i \cdot \psi_k(\mathbf{p}_i) \cdot \frac{1}{\rho(\mathbf{p}_i)} \right) \tag{4}$$

As the fixed size feature vector $\mathbf{v}$ is a function of the data points and their position, the feature vector can be expressed as a function of the following type:

$$\mathbf{v} : \mathbb{R}^i, \mathbb{R}^{i \times n} \mapsto \mathbb{R}^k \tag{5}$$

$\mathbf{v}$ describes the feature vector with a fixed size k as a reduction of an input of arbitrary size i for the data and $i \times n$ for the data's position for any i. Since the size of the feature vector $\mathbf{v}$ is fixed and does not depend on the input size i, the feature vector can be used as the input to a classical neural network architecture, such as a multi-layer feed-forward network, for an arbitrary input size i without the need to retrain the network.

V. Experiments

The method is tested with a dataset from a sensor based on multiple spatially resolved reflection spectroscopy (MSRRS, [10]). The data was measured in vivo, alongside a reference measurement of the carotenoid concentration in the skin on a scale ranging from 0 to 12. The measuring system, similar to the one described in [10], consists of several light emitters of different wavelengths, as well as several light detectors. The datasets for training and testing are entirely distinct, having measured a different group of test subjects using a different set of MSRRS-based sensors.

The measured spectroscopic data is well suited for the use of continuous kernels and continuous feature networks as proposed in section IV. This is because the MSRRS-based optical data is yielded in the shape of a relative brightness given for certain discrete wavelengths and certain discrete distances between light-emitter-detector pairs. These discrete wavelengths and distances are neither sampled at regular intervals nor always at the same exact wavelengths. Due to production inaccuracies for the sensors, slight differences in the wavelength of the emitters exist. However, the peak wavelengths are known for each sensor's emitters, and can thus be accurately supplied as the position data for the continuous feature layer. In addition, in this kind of spectral data, it is expected that the relevant data is encoded not in the shape of features to be detected, but in the position of the features (here the absorption wavelengths of the carotenoids), making the proposed method suitable.

To evaluate the method, a continuous feature network with a continuous feature layer containing 64 continuous kernels is used. Each kernel is made up of a SIREN network, containing three hidden layers of 48 nodes with sine nonlinearities each. The continuous feature layer is followed by a hidden feed-forward layer with 64 nodes, followed by an output layer with one output for the predicted carotenoid concentration. This network has approximately $320\mathrm{k}$ parameters.

For a comparison network, we use a multi-layer feed-forward neural network using a similar amount of parameters. This feed-forward network is supplied each emitter-detector pair as one node in the input layer, followed by a hidden layer of 256 nodes, followed by another hidden layer of 128 nodes, followed by an output layer with one output for the predicted carotenoid concentration, for a total of approximately $375\mathrm{k}$ parameters.

A convolution-based model was also investigated but it has proved unable to produce meaningful predictions of the carotenoid concentration in human skin and is thus omitted

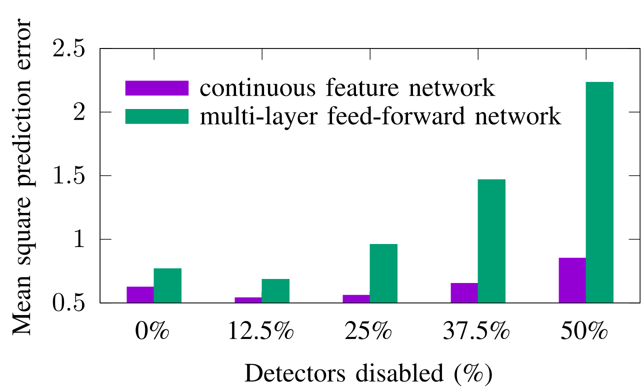

Fig. 2. The mean square prediction error (lower is better) of the continuous feature network and the multi-layer feed-forward network with the data of a different number of detectors withheld.

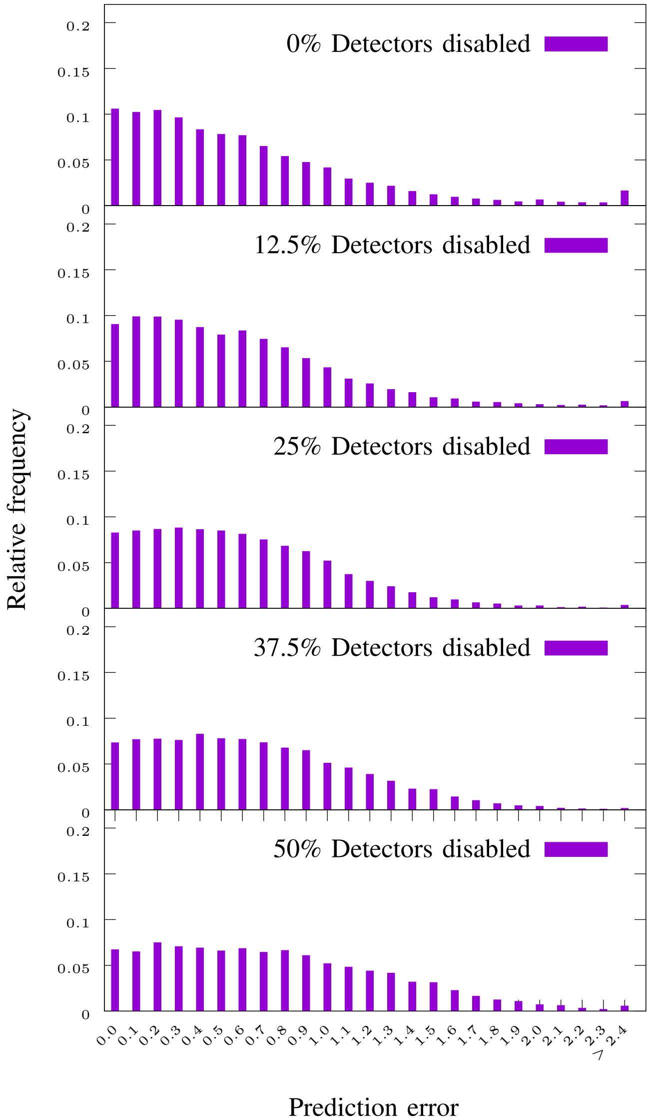

Fig. 3. A histogram of the prediction error of the continuous feature network at different numbers of light detectors whose data was withheld.

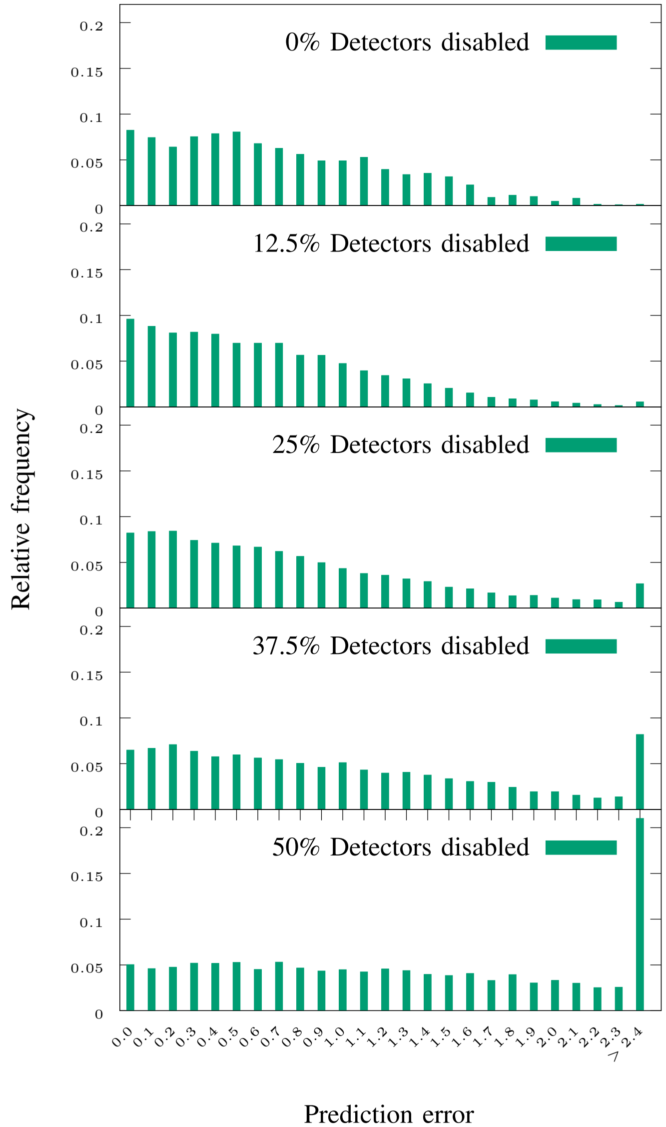

Fig. 4. A histogram of the prediction error of the multi-layer feed-forward network at different numbers of light detectors whose data was withheld.

from further analysis in this article.

Both networks were trained using the ADAM optimizer [11] and implemented using the LibTorch framework.

Figure 2 shows the accuracy of the proposed method compared to the accuracy of the comparison network. To show the ability of the continuous feature network to handle inconsistently sampled data, the prediction accuracy of the network was measured with the data of certain detectors withheld during inference. For each sample of data in the test set the detectors whose data was withheld were randomly picked, according to the number of detectors disabled. For the continuous feature layer, the missing data points were simply removed from the input vector. Due to the nature of the continuous feature network, it is capable of processing the shorter input vector without the need to retrain the model. For the multi-layer feed-forward network, the data was interpolated from the data of other detectors with a similar wavelength and emitter-detector distance. This is needed as the multi-layer feed-forward network is incapable of handling the shorter input vector without retraining. If no other data was available with a similar wavelength and distance, the value was set to 0 for the multi-layer feed-forward network. The results show that the continuous feature network outperforms the similarly-sized multi-layer feed-forward network for all investigated numbers of detectors whose data was withheld. The continuous feature network is able to achieve a mean square error of 19% lower compared to the multi-layer feed-forward network for the full set of input data. The improved prediction accuracy can be explained both by the high suitability of continuous feature networks for MSRRS data allowing an improved abstraction of the relationship between optical data and the reference carotenoid concentration in human skin, as well as because the continuous feature network is able to incorporate the actual measured wavelengths of the light emitters for each sensor as the position of the input data points. In addition, we see that the continuous feature network is able to give a stable prediction with more data missing compared to the multi-layer feed-forward network. This can also be seen in Figures 3 and 4. Figure 3 shows a histogram of the prediction error of the continuous feature network. In the different graphs, a different number of random light detectors were picked whose data was withheld from the continuous feature network. As the graph shows, while the amount of highly accurate precision lowers with more data being withheld, the amount of predictions with a large error (> 2.4) is not increasing significantly. This shows that the continuous feature network is capable of adapting to a lower amount of data being available to base its predictions on without the need for retraining. Figure 4 shows a histogram of the prediction error of the multi-layer feed-forward network. Similarly, the different graphs show the prediction error at different amounts of detectors whose data was withheld. In addition to a reduction in highly accurate predictions with more data withheld, the multi-layer feed-forward network quickly encounters an increase in predictions with a large error (> 2.4) once the amount of withheld data from disabled detectors increases to or above 25%. The slight increase of predictions with a large error occurring for the continuous feature network when no data is withheld is presumed to be due to an inaccurate input data point density estimation and will be subject to further investigation.

VI. Potential for Explainable AI

A side effect of the continuous feature layer is the resulting potential for explainable AI. As continuous kernels represent weights for each position in the measurement domain, we can deduct levels of importance of certain regions within the measurement domain from the encoded weights. The average of the absolute value of the weights over all kernels might be used as a measure of the importance of the data at certain points in the measurement domain. This may allow the use of the learned continuous kernels as an interpretable model [12]. However, as the magnitude of the input data at different positions in the measurement domain is not guaranteed to the normalized, any inferred importance from the kernel can be biased which will need to be accounted for. Similarly, the MLP being fed the latent feature vector will need to be considered when comparing the importance of the different kernels. Nonetheless, the usage of continuous feature layers as a tool for explainable AI is an interesting topic for further research.

VII. Conclusion and Future Work

This paper proposes the continuous feature network, a novel method to process irregularly and inconsistently sampled data with position-dependent features, such as optical or acoustic spectra. In addition, the continuous feature network is shown to outperform a comparable multi-layer feed-forward network with a 19% lower mean square error on predicting carotenoid concentration in human skin from optical multiple spatially resolved reflection spectroscopy data. This shows that the continuous feature network performed better at abstracting the relationship between the optical MSRRS data and the reference carotenoid concentration. Furthermore, this paper shows that the continuous feature network is capable of making stable predictions of carotenoid concentration in human skin with up to 50% of the data from the optical detectors withheld, while a comparable multi-layer feed-forward network exhibits a significant increase in predictions with a large error from 25% of the data withheld. Other potential use cases include similar types of data where samples may be irregular and features are position-dependent in the measurement domain, including other types of spectra, such as audio. Continuous feature networks also show potential for use as explainable AI and are worth studying further in this regard.

References

[1] A. R. T. Donders, G. J. van der Heijden, T. Stijnen, and K. G. Moons, "Review: A gentle introduction to imputation of missing values," *Journal of Clinical Epidemiology*, vol. 59, no. 10, pp. 1087–1091, 2006.

[2] P. K. Sharpe and R. J. Solly, "Dealing with missing values in neural network-based diagnostic systems," *Neural Computing & Applications*, vol. 3, no. 2, pp. 73–77, Jun 1995.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[4] D. W. Romero, A. Kuzina, E. J. Bekkers, J. M. Tomczak, and M. Hoogendoorn, "Ckconv: Continuous kernel convolution for sequential data," *CoRR*, vol. abs/2102.02611, 2021.

[5] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "Schnet: A continuous-filter convolutional neural network for modeling quantum interactions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, p. 992–1002.

[6] W. Wu, Z. Qi, and L. Fuxin, "Pointconv: Deep convolutional networks on 3d point clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9613–9622.

[7] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

[8] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 165–174.

[9] K. Genova, F. Cole, A. Sud, A. Sarna, and T. Funkhouser, "Local deep implicit functions for 3d shape," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2020, pp. 4856–4865.

[10] M. E. Darvin, B. Magnussen, J. Lademann, and W. Köcher, "Multiple spatially resolved reflection spectroscopy for in vivo determination of carotenoids in human skin and blood," *Laser Physics Letters*, vol. 13, no. 9, p. 095601, aug 2016.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[12] O. Biran and C. V. Cotton, "Explanation and justification in machine learning: A survey," in *IJCAI-17 Workshop on Explainable AI (XAI) Proceedings*, 2017.