YU.V. PUSTOVIT, YE.P. LYTVENIUK
Taras Shevchenko National University of Kyiv
(64/13, Volodymyrs'ka Str., Kyiv 01601, Ukraine; e-mail: jura.pustvit@gmail.com)

UDC 539

NEURAL-NETWORK-BASED METHODS FOR ARPES DATA PROCESSING (REVIEW ARTICLE)

In recent years, many developed upgrades of angle-resolved photoemission spectroscopy (ARPES) have significantly increased the amount of the obtained data. In this article, we briefly review the methods of processing of ARPES spectra with the use of convolutional neural networks (CNNs). In addition, we have made a short checkup of the potential application of CNNs that outperforms the existing methods or gives the possibility to achieve previously unachievable results.

K e y w o r d s: ARPES, convolutional neural network, machine learning.

1. Introduction

Angle-resolved photoemission spectroscopy (ARPES) is an experimental technique that is widely used to image directly surface and bulk electronic states. It is used in the studies of the electronic structure of various quantum materials ranging from strongly correlated states of matter to those exhibiting non-trivial topology [1–3]. The progress of the experimental tools has given rise to the improvement of the resolution and, hence, the information content of spectra. In recent years, a lot of upgrades of APRES such as time-resolved ARPES, spin-resolved APRES, spatially resolved ARPES, *etc.* have been developed [3]. The amount of the obtained ARPES data is increased nowadays, and the dimensionality becomes more complicated (from 1D to 3D), making the analysis of such spectra significantly more difficult [3]. In this case, various machine learning-based methods can be used to increase the efficiency of the data analysis. Machine learning (ML) methods have become one of the most widely used tools in different areas of science and technology [4, 5]. There are several reasons for why this happened. First of all, we mention the appearance of various open-source machine learning libraries such as Tensorflow [6], PyTorch [7], and sci-kit learn, which are user-friendly and give a possibility for the easy implementation of a vast range of different ML algorithms. Secondly, the extreme speed up in the training time through the graphics processing unit allows one to significantly decrease the time required for a model implementation. One of the best illustrations of the powerful cooperation between ML and fundamental physics is the Higgs Boson Machine Learning Challenge which was proposed to analyze the ATLAS experiment data in order to identify the Higgs boson [8]. In solid-state physics, ML methods showed their efficiency for the prediction of different material properties such as bulk and shear moduli, the existence of topological states or superconductivity, *etc.* [9–12]. In this article, we will make a short review of the existing methods of processing of ARPES spectra and determine the tasks that arise for ARPES and might be solved efficiently using various convolutional neural network-based methods.

C i t a t i o n: Pustovit Yu.V., Lytveniuk Ye.P. Neural-network-based methods for ARPES data processing (Review article). *Ukr. J. Phys.* **69**, No. 1, 53 (2024). https://doi.org/10.15407/ujpe69.1.53.

Ц и т у в а н н я: Пустовіт Ю.В., Литвенюк Є.П. Методи обробки APRES-спектрів на основі нейронних мереж (оглядова стаття). *Укр. фіз. журн.* **69**, №1, 53 (2024).

2. Convolutional Neural Networks

In the case of ARPES, the obtained data usually have 2 (image cuts of energy-momentum) or more (for example, results of time-resolved ARPES, when, in addition to energy and momentum, we have time) dimensions. One of the most widely used and efficient algorithms for the image processing is based on different types of convolutional neural networks. CNN is a neural network that is typically composed of three main types of layers: convolution, pooling, and fully connected [13, 14]. CNN is currently supposed to be one of the best algorithms that are used for the automated processing of images. As can be understood from the appellation, the core of the CNN architecture is a convolutional layer.

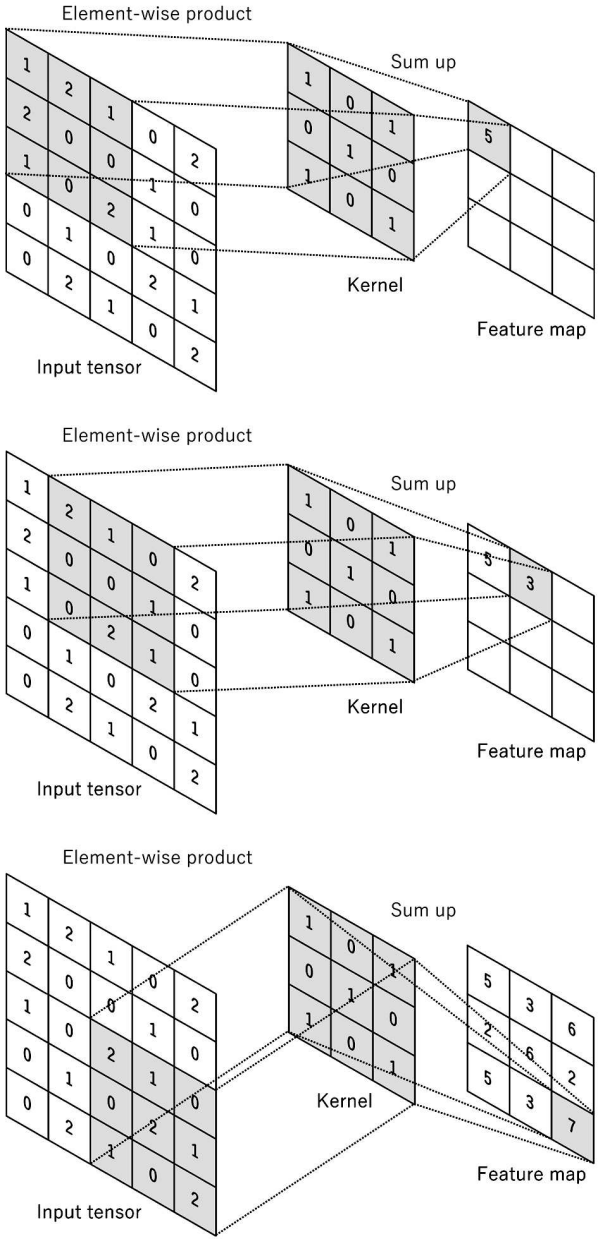

Fig. 1. An example of the convolution operation with a kernel size of (3,3) [13]

2.1. Convolutional layer

The convolutional layer is based on two operations – convolution and activation. Such layer consists of a set of kernels (also called filters), each of them performs the convolution operation with an input tensor (1-dimensional in the case of speech recognition, 2-dimensional for the image processing, or 3-dimensional for the video processing). In the image processing, convolution is the process of transforming an image by applying a kernel over each pixel and its local neighbors across the entire image and calculating an element-wise product between each element of the kernel and the corresponding pixel of the image (Fig. 1). The result of the application of the kernel to the image is called a feature map. During the neural network training process, the values of each kernel are changed to identify some kind of features from the input image. The advantage of this type of network is that it tries to find the correlation of neighboring pixels and learns to search for some kind of features in the input tensor, which is important for the following operation (processing with fully connected layers). In the case of predictive machine learning models, the performance is highly dependent on the feature engineering [12, 15]. In the case of CNN, the feature engineering is made with convolutional layers automatically. The hyperparameters of the convolutional layers are the kernel size (3 on 3 or 5 on 5, *etc.*), the number of kernels in the layer, and the activation function type.

2.2. Pooling layer

In CNNs, convolutional layers are often followed by a pooling layer. The reason to use the pooling is that the output of the convolutional layer determines the precise position of features in the input tensor. So, the rotation, shifting, or other changes in the feature position will result in another feature map. The pooling layer summarizes the features present in a

region of the feature map generated by a convolution layer. So, the summarized features are used for further operations instead of precisely positioned features generated by the convolution layer. This makes the model more robust to variations in the position of the features in the input image. One more reason to use pooling layers is to reduce the dimensions of the feature maps and to create a lower resolution version of the feature map, which contains the large or important structural elements, without the fine details and gives the possibility to reduce the number of parameters to learn [13]. In the case of pooling, the input tensor is divided into a set of regions that do not overlap, and such a region is replaced by one pixel. The pooling operation implementations might be different – max pooling (taking the maximal pixel value from the region), average pooling (taking the average pixel of all pixel values from the region), min pooling, *etc.* (Fig. 2).

2.3. Fully connected layers

Feature maps, which are obtained after convolutional and pooling layers are flattened then and used as the input on a fully connected neural network, which is used then for the classification or regression tasks [13, 14].

3. CNN-Based ARPES Spectra Processing Methods

To obtain the information from the ARPES spectra, they should be processed. There are a variety of tasks and approaches to such processing depending on various factors. We can generalize these different tasks into the two big families:

1. Determination of spectral parameters such as the electronic dispersion, real and imaginary parts of quasiparticle self-energy, gap size, *etc.* [1, 3].

2. Different image processing algorithms to extract different features and quick visual inspection, for example, the second-derivative method [16], curvature method [17], minimum gradient method [18], *etc.*

3.1. Determination of the electronic dispersion

Experimentally obtained ARPES spectra undergo the distortion and broadening in the energy and momentum dimensions due to the different interactions of electrons in the crystal (electron-electron and electron-phonon scatterings). In addition, the spectra are affected by various noises (from the inelastically scattered electrons, shot noise, *etc.*) and experimental setup apparatus function. Various methods are used to determine the electronic dispersion. It is worth to mention MDC and EDC analyses [32], and different image processing methods [16–18]. However, these methods have many different method-dependent constraints, and all of them tend to fail in complicated situations like:

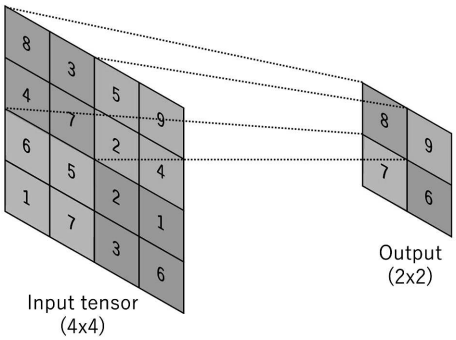

Fig. 2. The example of the max pooling operation [13]

1. Spectra with high levels of noise (low signal-to-noise ratio (SNR)).

2. Multiband systems, when two or more bands are in the vicinity of each other, the intersection of different bands, the intensity of one band is much lower than those of the others.

One of the successful examples of applying neural networks for the ARPES spectra processing is the determination of the electronic band dispersion [19]. In this case, the electronic dispersion determination has been regarded as finding an appropriate inverse transformation that can best recover the original images – which can be formulated as solving an inverse problem. To solve this problem, it has been proposed to use a CNN, which provides a great performance in improving the image qualities by solving a series of inverse problems such as the super-resolution, denoising, and patching. The idea of neural network training can be seen in the example – we have the training data consisting of input (in our case, simulated experiment data Y) and the desired output ("dressed") electron dispersion (simulated energy band X). During the forward pass, each filter is convolved across all dimensions of the input volume (width, height, depth), computing the dot product between the filter entries and the input, producing a feature map of

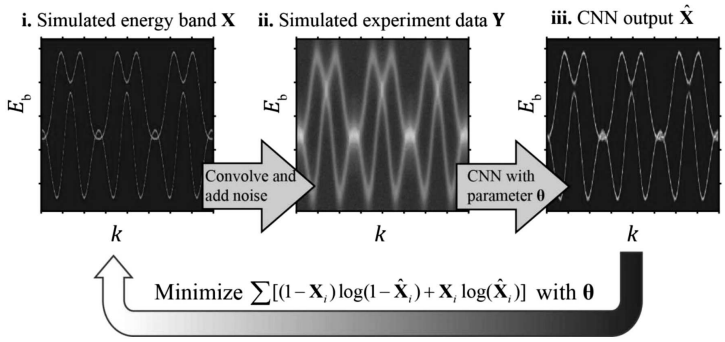

Fig. 3. The example of the simulated desired data (X), simulated experiment data Y, and neural network output [19]

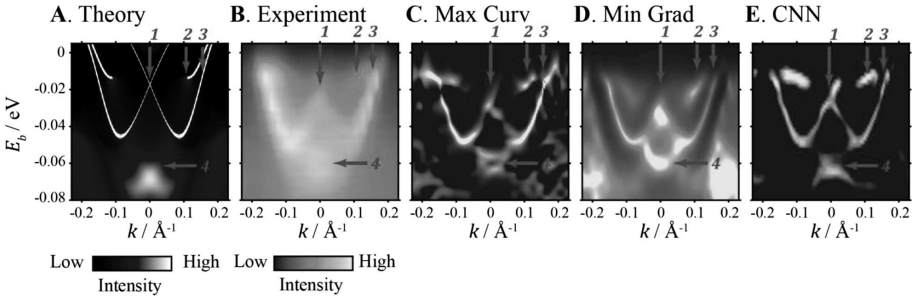

Fig. 4. The example of the experimental ARPES spectra processing within different methods [19]

that filter. These are then taken through an activation function, which decides whether a certain feature is present at a given location in the image. During the training process, the kernels (filters) of such a network are changed to minimize the loss function (difference between desired and obtained outputs). This process is repeated until the minimum loss function is achieved or some number of epochs have been passed. The results of the spectra processing using such a neural network and with other methods are shown in Fig. 4. It can be seen that the CNN is the most precise method for the electronic dispersion determination, which enables a detailed extraction of the characteristic features [19].

3.2. Noise removal

One more practical application of neural networks is the denoising of ARPES spectra. The authors of work [20] proposed a neural network that consists of 20 convolutional layers, each layer of the convolutional neural network has a filter number of 64 and a filter size of 3. The training data set consists of pairs of original and generated data. The desired output is the spectra with a high signal-to-noise ratio (can be used readily available ARPES data), and the input is the corresponding spectra with noise addition (generated spectra with a low signal-to-noise ratio).

As can be seen in Fig. 5, *b*, unlike the Gaussian smoothing, the denoising neural network efficiently removes noise, and simultaneously preserves the intrinsic band structure. The reason for this is that the size of the characteristic feature of the band structure is much greater than the data pixel size (any feature occupies much more than one pixel). Such a neural network adapts to get more information from the noisy input than just the pixel-wise sum using the a priori information extracted from the training set. So, if the information at a pixel is corrupted with noise, the value can be recovered, from the most statistically probable value inferred from adjacent pixel values [20].

3.3. Autoencoder-based noise reduction and feature extraction

The above-mentioned methods are used for the band tracing and denoise simultaneously the ARPES spec-

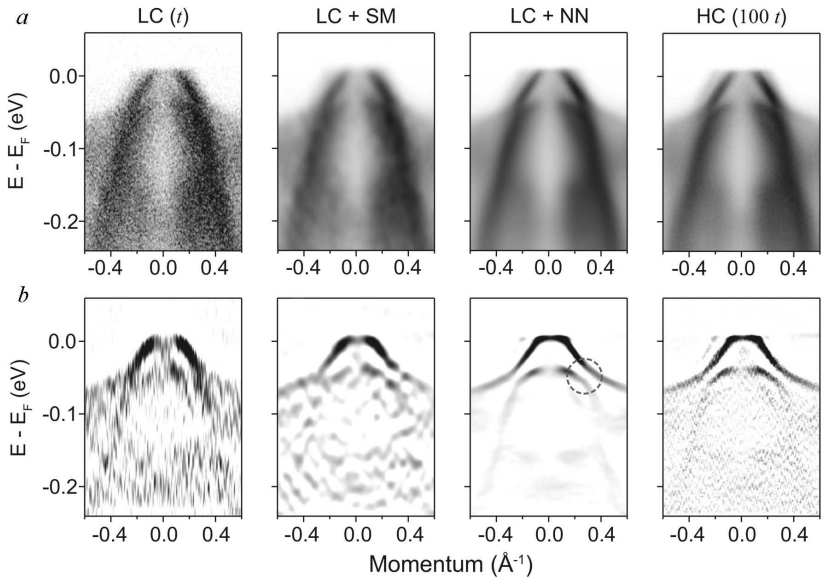

Fig. 5. Denoising results of the neural network proposed in [20]. (*a*) FeSe ARPES data long the M-Γ-M cut and denoising results. LC and HC denote low-count (low SNR) and high-count data (high SNR). t denotes a unit acquisition time. The low- and high-count data are acquired for t and 100 t, respectively. SM and NN denote the Gaussian smoothing and denoising neural network, respectively. (*b*) Corresponding second-derivative results of (*a*) [20]

tra. To trace the bands and to denoise spectra simultaneously, a special convolutional neural network was proposed [21] – an autoencoder [22, 23]. Such a network aims to learn data encodings (lower-dimensional representation(encoding) for higher-dimensional data and have numerous practical applications such as the denoising, colorization, feature-level arithmetic, detection, tracking, and segmentation [24, 25]. To train the denoising autoencoder, images with noise are used as the input, and noise-free images are used as the output [24, 25]. The autoencoder consists of 3 parts:

1. Encoder (compresses the input data).

2. Bottleneck (contains the compressed knowledge representations).

3. Decoder (reconstructs data from their encoded form).

The first is the encoder, which transforms the input distribution into a low-dimensional tensor (latent representation vector). Since the latent vector is of a low dimension, the encoder is forced to learn only the most important features of the input data. The second part is the decoder, which tries to recover the input from the latent vector. As the latent vector is the low-dimensional compressed representation of the input distribution, the recovered output can only approximate the input. The goal of the decoder is to make the output as close as possible to the input. One of the most commonly used cases of such neural networks is the image denoising. Unlike "classical" denoising methods, autoencoders extract the image from the noisy input data rather than search for noise [21, 23–25].

It has been shown that such a convolutional variational autoencoder (VAE) network can simultaneously perform the denoising and feature extraction from ARPES dispersion maps. The efficiency of the VAE is a result of the combination of convolutional layers (which are useful in the denoising) and a bottleneck (which enables the feature extraction). Autoencoder-like topology is used to extract only important features for the image reconstruction. It has been shown that the proposed neural network can just denoise the image or denoise the image and extract features simultaneously. For the spectra denoising, the networks are trained using generated data with many-body renormalizations and broadening. To generate the data, the spectral function has been used. It includes the effects of the interaction of fermionic quasiparticles with a collective bosonic mode. The spectral function was convolved the spectral function with a Gaussian resolution function and multiplied by a Fermi function. To obtain a variety of images neces-

sary for the network training, most parameters randomly varied in the range of values near the real parameters of the Bi2212 [21].

As the input, noisy versions of the generated spectra were used, and the desired output was presented by generated spectra without noise. The size of the whole set was 6000 pairs (5000 – training set and 1000 – validation set). The denoising of the Bi2212 spectra with the proposed network has revealed that the obtained energy distribution curves (EDC) follow the underlying line shape quite accurately and show almost no noise.

The more complicated model of the training data generation was used for the spectra denoising with the feature extraction. It contains a bigger amount of parameters (15 different parameters in contrast to 9 in the previous case). The size of the whole set was 9000 pairs (7000 – training set and 2000 – validation set). The desired output of VAE was the generated spectra, and the input was the corresponding broadened, noisy, and anisotropic versions of these spectra. Processing the experimentally obtained spectra of 1T-TiSe2 has shown that this network could resolve the dispersion of the top and bottom branches of the Bogoliubov-like bands around the M point of the Brillouin zone [21].

There are two peculiarities of the above-mentioned neural networks that show the powers of such an approach:

1. Though the training networks have been used for generated datasets, both networks gave results on experimentally obtained spectra, which are better than the results of the conventional methods.

2. Both networks are used efficiently for solving the various inverse problems (from noisy data to denoised images; from the distorted electronic dispersion to the 'bare' one).

4. Train-Set-Free Denoising

The absence of the experimentally obtained spectra without the noise is a complication in the case of different denoising neural network methods. To overcome this limitation, several approaches are used. First, it is the usage of generated spectra as the desired output, as well as the corresponding spectra with additional noise as the input [21]. Another approach is based on usage of experimentally obtained spectra with a high signal-to-noise ratio as the desired output. In this case, the desired result is such a spectrum, and the input is this spectrum with adding a synthetic noise [20]. However, these approaches lack flexibility and tend to fail with spectra that differ dramatically from the one in the training set.

The training-set-free method was proposed to avoid the collection of training sets and achieve higher adaptability and flexibility. This method is based on the decomposition of the spectra into two parts: correlated part and a noise. Such an approach is valid for the ARPES spectra in which nearby pixels are highly correlated. Two independent neural networks have parametrized the clean spectra and the noise. The correlated part has been parametrized as the output of an encoder-decoder-type convolutional neural network. The encoder part maps the input image to the latent space, the decoder plays the opposite role and maps the data from the latent space to the input space. The input of this network can be somewhat arbitrary, and the only requirement might be the reflection of self-correlation. These two neural networks first guess a superposition of the clean spectrum and noise via the forward propagation. Then the parameters of the neural network are updated concerning the loss function through the backpropagation (BP). The main advantage of such a method is that it can be easily extended to the three-dimensional case, without requiring a large amount of expensive high-quality training data and a time- or/and resource-consuming training process [26].

Such a method has been used to extract a grid structure in ARPES spectra due to the metal mesh in front of the analyzer. In this case, the spectrum is decomposed into three textures: the clean spectra, the grid structure, and the noise. To parametrize the observed data, two convolutional neural networks have been used. One network was used to extract the "clean" spectra, and the other to determine the grid texture and noise. These two networks allow separating the signal of the energy bands from the grid, even when the grid width and the energy bandwidth are comparable (Fig. 6). The main advantage of the deep-learning method is the possibility of extracting the complete intrinsic energy band signal using the local data correlation to prevent losing the critical information [27].

5. Perspectives

Besides the tasks mentioned above, there are a bunch of different problems that can be solved more effi-

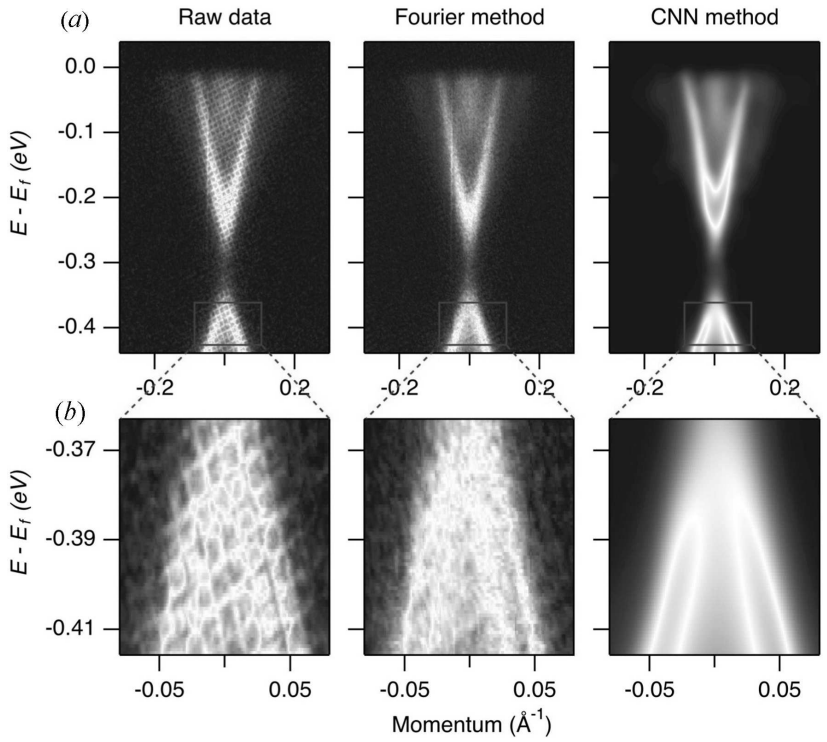

Fig. 6. Comparison of the degridded and denoising results using different methods. $MnBi_2Te_4$ ARPES data along the cut through Γ and the degridded results based on the Fourier filtering method and the deep-learning method proposed in [26] (*a*). (*b*) Close-up showing of the area in the box region indicated in panel (*a*) [26]

ciently or solved using neural networks only. We will determine such issues and propose a neural network-based approach for solving them.

5.1. Physical model determination

To have a reliable approach to determine the underlying physical model from ARPES spectra, a classifier based on CNN can be used. Such networks are used for a wide variety of various tasks such as the prediction of Alzheimer's disease from the magnetic resonance imaging [28], the detection of common rice plant anomalies [29], and neutrino event classification [30]. However, the topology of such networks does not vary greatly. They all consist of repeating convolution layers followed by pooling layers. The final pooling layer output is flattened out, then transformed into a one-dimensional array, and fed to the fully connected layers that predict the output. The goal is to train such a neural network to assign the generated spectra to the corresponding model that has been used to generate spectra. Considering the ability of neural networks to generalize on generated data can achieve significant classification results for the experimentally obtained spectra.

As the example of such a problem, we can consider the physical mechanism of the so-called "pseudogap" in copper-based superconductors. The opening of a gap or just a depletion of the electronic density of states at the Fermi level has been observed by different experimental techniques, but ARPES is the most direct tool to access the electronic density of states at the Fermi level. The physical mechanism of the pseudogap appearance in copper-based superconductors remains unclear. It is supposed that the pseudogap in cuprates is a complex phenomenon that includes at least three different "intertwined" orders: spin and charge density waves (similar to the 2D CDW compounds) and preformed pairs, which appear in different parts of the phase diagram [31]. A major issue arises from the lack of a dependable criterion to differentiate between various types of gaps in ARPES data. The input of the network in this case is the

image of a certain dimensionality (spectra generated using a certain theoretical model), the output integer number or vector with "one-hot encoding" that corresponds to the predicted class (corresponding theoretical model). If the differences between gaps can be established from the spectra – the neural network will learn how to do it.

5.2. ARPES-spectra parameter extraction

The conventional model to describe ARPES spectra is a one-particle spectral function, which is the imaginary part of the Green's function for one-electron excitations (quasiparticles):

$$A(\omega, k) = \frac{\Sigma''(\omega)}{(\omega - \varepsilon(k) - \Sigma'(\omega))^2 + \Sigma''(\omega)^2},$$

where $\Sigma'(\omega)$, $\Sigma''(\omega)$ – is the real and the imaginary parts of the quasiparticle self-energy, which reflects all the interaction of electrons in the crystal, $\varepsilon(k)$ – "bare" electron dispersion. Thus, not only the structure of one-electron bands, but also the structure of basic interactions in the electronic system might be expected to reflect in the ARPES spectrum. To determine the quasiparticle self-energy, the MDC and EDC-analyses are used [1, 32–34] In the case of the MDC analysis, the spectrum is divided into a set of momentum distribution curves at fixed energies. Each of these curves is fitted using the Lorenz function, which gives the possibility to extract the imaginary part of the self-energy that corresponds to the full-width half maximum (FWHM) of the Lorenzian, and the maximum of the curve is determined by $\omega - \varepsilon(k) - \Sigma'(\omega)$. From whence, we can determine the real part of the self-energy. The MDC analysis can be used only for not too flat bands (does not work near and on the top/bottom of bands), the dependence of the self-energy on the momentum can be neglected (usually valid) [32]. In the case of the EDC-analysis, the spectrum is divided into a set of the energy distribution curves at fixed energies. Each of these curves is fitted using a model (there is no conventional model as in the case of the MDC analysis). From this model, we can obtain the self-energy and the electronic dispersion. However, there are also a lot of restrictions. For example, EDC does not work for a too steep dispersion, suppose the self-energy does not depend on the energy (that is not valid in most cases). In the EDC analysis, the noise of inelastically scattered electrons becomes significant, because it is not constant and dependent on the energy and should be also considered in the model used for the fitting.

So, we can conclude that both methods have restrictions and involve the correlation only between pixels in one dimension. Another disadvantage is the difficulty of the processing automation (which is important for the processing of large amounts of data).

The CNN-based method can help to overcome the limitations of the existing methods:

1. Convolutional layers search for a correlation between pixels in all directions and treat spectra as the whole, not as the set of almost independent slices.

2. All spectral parameters are determined almost instantly, and a large amount of data can be processed using such a method.

The tradeoff is the lack of versatility – the network works only for one particular physical model (defined by the training set generator) and the set of parameters (defined by the output layer topology). However, there is a simple way to make this approach more universal – using more than one neural network. Each of the networks corresponds to a different model with a different set of parameters. As the differences between networks are not essential, the transfer learning [35] reduces drastically the time for the neural network training.

5.3. Time-resolved ARPES (trARPES) analysis

TrARPES is the experimental technique that gives the possibility to investigate not only the electronic band structure, but also the momentum-resolved electronic dynamics. The obtained data give the possibility to get non-equilibrium properties of a variety of different materials. TrARPES is performed in a pump-probe configuration using an ultrafast laser system, the pulse excites a sample out of equilibrium, and the probe pulse is used for measuring the system from the transient system. The ARPES spectra are recorded as a function of the pump-probe time delay [3, 36]. So, the trARPES gives the possibility to observe the additional time dimension. However, during the analysis of such data, each spectrum is handled separately using the EDC and MDC analyses or other methods, reducing the 3-dimensional task to a set of one-dimensional ones [37–39]. Then, by comparing the obtained results for different delay times,

we determine the differences and extract information about the systems under investigation. Such an approach is not reliable and can not be used to extract non-evident features and peculiarities.

New methods should be used to exploit all benefits of this experimental technique for tracking both spatial and time correlations between the spectra. The approaches for the video classification tasks are considered in [40–42]. In this case, more information is provided to the recognition task by adding a temporal component through which plenty of other information can be used. In addition, this task is much more computationally demanding even for processing the short videos, since each video might contain hundreds to thousands of frames, not all of which are useful, and the training of such networks can not be parallelized (GPU will not increase the training speed). However, a combination of trARPES and different types of recurrent CNNs (CNN + RNN [40] or CNN + LSTM (long short-term memory) [43]) gives the possibility to verify different models based not only on the peculiarities of the electronic band structure, but also adding the temporal component through which other information can be additionally used.

An example of such a problem is the classification of the physical processes and identifying the relaxation mechanism after the absorption of pump photons. The situation when the absorption of pump photons results in optical dipole transitions from occupied to unoccupied electronic states with the subsequent relaxation due to the intrinsic scattering process is too simplified. When the material is in an ordered state, the optical excitation can perturb or destroy the order parameter. The excitation can launch a coherent mode such as a photon via the Raman process or electrons can be "dressed" by the periodic structure of the pump fields and form Floquet–Bloch states. Of course, each of these mechanisms does not exclude another one. So, complicated combinations of different mechanisms might exist. Even from the time-resolved spectra, it is a challenge to classify the arising physical process and identify the existing relaxation mechanisms [3].

5.4. Predicting critical temperature of the high-temperature superconductors

Now, plenty of machine learning methods are used to predict the properties of materials such as topological states, band gaps, bulk and shear moduli, superconductivity, and others. Such machine-learning-based models use different parameters such as the composition-weighted average of a number of unfilled electrons, the maximum value of melting temperature, and the standard deviation of the number of valence electrons taken over the element value of the compound [12, 44].

In the case of different predictive models, the critical step in the model creation is the feature engineering. The feature engineering (extraction) is the batch of methods that select and/or combine variables into features, effectively reducing the amount of data that must be processed, while still wholly and accurately describing the original data set. Such a step is useful to reduce the number of resources needed for the model training and to increase the efficiency of the model. However, not all important features can be constructed from the proposed variables [12, 44].

For example, there are a huge number of different compounds, where the appearance/disappearance of the superconducting state or essential changes of the critical temperature are accompanied by complicated changes in the electronic band structure [45, 46]. Some empirical correlations between superconductivity and the electronic band structure have been established [45]. But not all important features can be established from a 2D slice or the set of 2D slices of the 3D Brillouin zone almost "by eye".

The Fermi surface tomography enables performing a 3-D mapping of the Fermi surface within a short time interval with very high resolution [47]. CNN allows the building of predictive models for the Fermi surface top with the existence of high-temperature superconductivity or with drastic changes in the critical temperature. Their convolutional layers learn to find peculiarities of the electronic band structure, which are crucial for the high-temperature superconductivity, and fully connected layers predict the critical temperature using extracted features [4, 5].

The problem with such an approach is the absence of a standardized database due to the lack of the experimentally obtained ARPES spectra for different compounds with corresponding critical temperatures. If this database will be created, after performing the computationally inexpensive neural network training, using the Open Quantum Materials Database (OQMD) [48] or another DFT calculation database search for the high-temperature superconductors among different compounds.

Another possible approach (that is easier to interpret) is to use convolutional and pooling layers just for the feature extraction, and the extracted features will be used for other machine learning models such as the random forest one [49, 50].

6. Discussion

We have reviewed the pros of such CNN-based models. However, we should say a few words about some cons:

1. There is some skepticism about the usage of neural networks for different scientific tasks due to a lack of interpretability. In most cases, the neural network is regarded as a "black box" with an unknown decision-making mechanism. So, even in the case of creating a powerful predicting model, we will not be able to determine new physical laws. However, in recent years, the number of articles devoted to the interpretability of neural networks has grown exponentially [51, 52, 52]. A lot of methods aimed to increase the transparency of the model such as the feature analysis techniques, different model inspections, saliency methods, *etc.* have been proposed. Almost all these methods are time and computationally expensive, not well worked out yet, and are not universal. Although it is unlikely that such approaches will make neural network-based models such transparent and clear as the simple linear models, the idea of "black box" tends to be completely incorrect [51, 52].

2. The convolutional layers give the possibility to find the dependences between the pixel and its local neighbors, but they tend to fail in the case of long-range dependences. The attention mechanism [54] has been regarded as an advanced technique to capture long-range dependences. So, some additional complications of the neural network give the possibility to overcome the limitations of the kernels to determine and consider long-range dependences between pixels [55–57].

An alternative popular approach for the computer vision tasks is the usage of visual transformers (ViT) [58]. Such models employ a transformer-like architecture [54], which has been widely recognized as a powerful tool for solving the natural language processing tasks [59, 60]. The performance of ViT has outperformed CNNs on multiple benchmark tests [60, 61]. However, such improvements are achieved mostly for large datasets (more than several millions of examples) and pre-trained models. So, in the case of ARPES spectra processing, where the datasets are much smaller [19–21], ViT does not tend to have better results than the CNN.

7. Conclusion

We have reviewed the existing CNN-based methods for the ARPES spectra processing, that are used for the noise reduction and/or the electronic dispersion determination. Such methods outperform the "classical" ARPES processing methods such as the second-derivative, minimum curvature, maximum gradient, Gaussian denoising, and other ones. There are numerous reasons for the efficiency of such networks. First of all, we mention the ability to generalize. So, the use of relatively small training datasets (less than 10000 examples) of generated data networks allows one to efficiently process the experimentally obtained spectra. Second, due to the adaptability and flexibility, almost the same networks can be used for significantly different tasks simply by changing the training dataset. Third, the milestone of all reviewed methods is the ability to learn and use the complicated correlation between neighboring pixels using the convolutional layers. Such correlation allows the denoising or degridding of spectra using the decomposition of spectra into uncorrelated and correlated parts. In this case, CNN is used to approximate the correlated part, which corresponds to "pure" spectra.

However, besides the existing methods, there is a wide range of tasks that can be resolved using CNNs. We have outlined some of them such as:

1. trARPES spectra processing using the different types of recurrent CNNs;

2. physical model classification using a CNN-based classifier;

3. automated ARPES spectral parameter extraction using CNNs;

4. creating predicting models for predicting the critical temperature and/or existence of the superconductivity of the high-temperature superconductors based on the electronic band structure.

We acknowledge discussions with A.A. Kordyuk and A.M. Kutsyk. This work was partly supported by the National Research Foundation of Ukraine through the project 2020.02/0408 and by grant No. 22BF07-03 from the Ministry of Education and Science of Ukraine.

1. A.A. Kordyuk. ARPES experiment in fermiology of quasi-2D metals (Review Article). *Low Temp. Phys.* **40**, 286 (2014).
2. A. Damascelli, Z. Hussain, Z.-X. Shen. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473 (2003).
3. J.A. Sobota, Yu He, Z.-X. Shen. Angle-resolved photoemission studies of quantum materials. *Rev. Mod. Phys.* **93**, 025006 (2021).
4. Shi Dong, Ping Wang, Khushnood Abbas. A survey on deep learning and its applications. *Comp. Sci. Rev.* **40**, 100379 (2021).
5. S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. Presa Reyes, M.-L. Shyu, S.-C. Chen, S. Iyengar. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput. Surv.* **51**, 1 (2018).
6. A. Mart, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. Gordon Murray *et al.* TensorFlow: A system for large-scale machine learning. arxiv.org/abs/1605.08695.
7. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, Zach M. Raison *et al.* PyTorch: An imperative style, High-performance deep learning library. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, Canada, December 8–14, 2019.*
8. C. Adam-Bourdarios, G. Cowan, C. Germain-Renaud, I. Guyon, B. Kégl, D. Rousseau. The higgs machine learning challenge. *J. Phys. Conf. Ser.* **664**, 072015 (2015).
9. J. Dean, M. Scheffler, A.R.P. Thomas, S.V. Barabash, R. Bhowmik, T. Bazhirov. The higgs machine learning challenge. arXiv: 2112.00239 [cond-mat.mtrl-sci].
10. P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Mollo, M. Zeller, S.A. Friedler, J. Schrier, A.J. Norquist. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73 (2016).
11. S.R. Xie, Y. Quan, A.C. Hire, B. Deng, J.M. DeStefano, I. Salinas, U.S. Shah, L. Fanfarillo, J. Lim, J. Kim, G.R. Stewart, J.J. Hamlin, P.J. Hirschfeld, R.G. Hennig. Machine learning of superconducting critical temperature from Eliashberg theory. *Npj Comput. Mater.* **8**, 14 (2022).
12. J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 1 (2019).
13. R. Yamashita, M. Nishio, R.K.G. Do, K. Togashi. Convolutional neural networks: An overview and application in radiology. *Insights into Imaging* **9**, 611 (2018).
14. K. O'Shea, R. Nash. An introduction to convolutional neural networks. arXiv:1511.08458.[cs.NE].
15. I. Guyon, A. Elisseeff. *Feature Extraction: Foundations and Applications.* Edited by I. Guyon, M. Nikravesh, S. Gunn, L.A. Zadeh (Springer, 2013) [ISBN: 978-3-540-35487-1].
16. R.C. Gonzalez, R.E. Woods. *Digital Image Processing* 3rd edition (Prentice Hall, 2008) [ISBN: 978-0131687288].
17. P. Zhang, P. Richard, T. Qian,Y.-M. Xu,X. Dai, H.Ding. A precise method for visualizing dispersive features in image plots. *Rev. Sci. Instrum.* **82**, 043712 (2011).
18. Yu He, Y. Wang, Z.-X. Shen.Visualizing dispersive features in 2D image via minimum gradient method. *Rev. Sci. Instrum.* **88**, 073903 (2017).
19. H. Peng, X. Gao, Yu He, Y. Li, Y. Ji, Ch. Liu, S.A. Ekahana, D. Pei, Z. Liu, Z. Shen, Yu. Chen. Super resolution convolutional neural network for feature extraction in spectroscopic data. *Rev. Sci. Instrum.* **91**, 033905 (2020).
20. Yo. Kim, D. Oh, S. Huh, D. Song, S. Jeong, Ju. Kwon, M. Kim, D. Kim, H. Ryu, J. Jung, W. Kyung, B. Sohn, S. Lee, J. Hyun, Ye. Lee *et al.* Deep learning-based statistical noise reduction for multidimensional spectral data. *Rev. Sci. Instrum.* **92**, 073901 (2021).
21. F. Restrepo, Ju. Zhao, U. Chatterjee. Deep learning-based statistical noise reduction for multidimensional spectral data. *Rev. Sci. Instrum.* **93**, 065106 (2022).
22. G.E. Hinton, R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science* **331**, 504 (2006).
23. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**, 3371 (2010).
24. R. Atienza. *Advanced Deep Learning with Keras: Apply Deep Learning Techniques, Autoencoders, GANs, Variational Autoencoders, Deep Reinforcement Learning, Policy Gradients, and More* (Packt Publishing, 2018) [ISBN: 978-1788629416].
25. R. Atienza. *Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, Deep RL, Unsupervised Learning, Object Detection and Segmentation, and More* (Packt Publishing, 2020) [ISBN: 978-1838825720].
26. D. Huang, J. Liu, T. Qian, Y.F. Yang. Spectroscopic data de-noising via training-set-free deep learning method. *Sci. China: Phys. Mech. Astron.* **66**, 267011 (2023).
27. J. Liu,D. Huang, Y.F. Yang, T. Qian. Removing grid structure in angle-resolved photoemission spectra via deep learning method. *Phys. Rev. B* **107**, 165106 (2023).
28. Y. Huang, J. Xu, Y. Zhou, T. Tong, X. Zhuang. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. *Front. Neurosci.* **13**, (2019).
29. R. Atole, D. Park. A multiclass deep convolutional neural network classifier for detection of common rice plant anomalies. *J. Adv. Comput. Sci. App.* **9**, (2018).
30. A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M.D. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, P. Vahle. A convolutional neural network neutrino event classifier. *J. Instrum.* **11**, P09001 (2016).
31. A.A. Kordyuk. Pseudogap from ARPES experiment: Three gaps in cuprates and topological superconductivity (Review Article). *Low Temp. Phys.* **41**, 417 (2015).

32. T. Valla, A.V. Fedorov, P.D. Johnson, B.O. Wells, S.L. Hulbert, Q. Li, G.D. Gu, N. Koshizuka. Evidence for quantum critical behavior in the optimally doped cuprate. *Science* **285**, 2110 (1999).
33. Z.-X. Shen, J. R. Schrieffer. Momentum, temperature, and doping dependence of photoemission lineshape and implications for the nature of the pairing potential in high-T_c superconducting materials. *Phys. Rev. Lett.* **78**, 1771 (1997).
34. P.A. Casey, J.D. Koralek, N.C. Plumb, D.S. Dessau, P.W. Anderson. Accurate theoretical fits to laser-excited photoemission spectra in the normal phase of high-temperature superconductors. *Nat. Phys.* **4**, 210 (2008).
35. K. Weiss, T,M. Khoshgoftaar, D. Wang. A survey of transfer learning. *J. Big. Data* **3**, 9 (2016).
36. A.F. Kemper, M.A. Sentef, B. Moritz, T.P. Devereaux, J.K. Freericks. Review of the teoretical description of time-resolved angle-resolved photoemission spectroscopy in electron-phonon mediated superconductors. *Ann. Phys.* **529**, 1600235 (2017).
37. W. Zhang, C.L. Smallwood, C. Jozwiak, T.L. Miller, Y. Yoshida, H. Eisaki, D.-H. Lee, A. Lanzara. Signatures of superconductivity and pseudogap formation in nonequilibrium nodal quasiparticles revealed by ultrafast angle-resolved photoemission. *Phys. Rev. B* **88**, 245132 (2013).
38. C.L. Smallwood, W. Zhang, T.L. Miller, C. Jozwiak, C.H. Eisaki, D.-H. Lee, A. Lanzara.Time- and momentum-resolved gap dynamics in $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. B* **89**, 115126 (2014).
39. R. Cortés, L. Rettig, Y. Yoshida, H. Eisaki, M. Wolf, U. Bovensiepen. Momentum-resolved ultrafast electron dynamics in superconducting $Bi_2Sr_2CaCu_2\mathbf{O}_{8+\delta}$. *Phys. Rev. Lett.* **107**, 097002 (2011).
40. E. Apostolidis, E. Adamantidou, A.I. Metsai, V. Mezaris, I. Patras. Video summarization using deep neural networks: A survey. arXiv:2101.06072.[cs.CV].
41. J.Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici. Beyond short snippets: Deep networks for video classification. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 7–12, 2015.*
42. C. Orozco, E. Xamena, M. Buemi, J. Berlles. Human action recognition in videos using a robust CNN LSTM approach. *Ciencia y Tecnología* 21 (2020).
43. A. Agga, A. Abbou, M. Labbadi, Y. El Houm, I. Hammou Ou Ali. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research* **208**, 107908 (2022).
44. V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi. Machine learning modeling of superconducting critical temperature. *Npj Comput. Mater.* **4**, 29 (2018).
45. A.A. Kordyuk. Electronic band structure of optimal superconductors: From cuprates to ferropnictides and back again. *Low Temp. Phys.* **44**, 477 (2018).
46. Yu.V. Pustovit, A.A. Kordyuk. Metamorphoses of electronic structure of FeSe-based superconductors (Review article). *Low Temp. Phys.* **42**, 995 (2016).
47. S. Borisenko, A. Fedorov, A. Kuibarov, M. Bianchi, V. Bezguba, P. Majchrzak, P. Hofmann, P. Baumgärtel, V. Voroshnin, Y. Kushnirenko, J. Sánchez-Barriga, A. Varykhalov, R. Ovsyannikov, I. Morozov, S. Aswartham. Fermi surface tomography. *Nat. Commun.* **13**, 4132 (2022).
48. S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton. The open quantum materials database (OQMD): Assessing the accuracy of DFT formation energies. *Npj Comput. Mater.* **1**, 15010 (2015).
49. A. Wang, Y. Wang, Y. Chen. Hyperspectral image classification based on convolutional neural network and random forest. *Remote Sens. Lett.* **10**, 1086 (2019).
50. G. Xu, M. Liu, Z. Jiang, D. Söffker, W. Shen. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors* **19**, 1088 (2019).
51. F.-L. Fan, J. Xiong, M. Li, G. Wang. On interpretability of artificial neural networks: A survey. *IEEE trans. Radiat. Plasma Med. Sci.* **5**, 741 (2020).
52. Y. Zhang, P. Tiňo, A. Leonardis, K. Tang. A survey on neural network interpretability. *IEEE Trans. Emerg. Top. Comput. Intell.* **5**, 726 (2021).
53. Q. Zhang, Y.N. Wu, C.-H. Zhu. Interpretable convolutional neural networks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 18–23, 2018.*
54. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin. Attention is all you need. In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems Long Beach, California, USA December 4–9, 2017.*
55. Chunwei Tian, Yong Xu, Zuoyong Li, Wangmeng Zuo, Lunke Fei, Hong Liu. Attention-guided CNN for image denoising. *Neural Netw.* **124**, 117 (2020).
56. Z. Baozhou, P. Hofstee, J. Lee, Z. Al-Ars. An attention module for convolutional neural networks. arXiv: 2108.08205.[cs.CV].
57. M.-I. Georgescu, R.T. Ionescu, Andreea-Iuliana Miron, Olivian Savencu, Nicolae-Catalin Ristea, Nicolae Verga, Fahad Shahbaz Khan. Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution. arXiv:2204.04218.[eess.IV].
58. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv:2010.11929.[cs.CV].
59. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.[cs.CL].
60. C. Li, C. Zhang. Toward a Deeper understanding: RetNet viewed through convolution. arXiv:2309.05375.[cs.CL].

61. O. Uparkar, J. Bharti, R.K. Pateriya, R. Kumar Gupta, A. Sharma. Vision transformer outperforms deep convolutional neural network-based model in classifying X-ray images. *Procedia Comput. Sci.* **218**, 2338 (2023).

Received 29.10.23

Ю.В. Пустовіт, Є.П. Литвенюк

МЕТОДИ ОБРОБКИ
ARPES-СПЕКТРІВ НА ОСНОВІ
НЕЙРОННИХ МЕРЕЖ (ОГЛЯДОВА СТАТТЯ)

Вдосконалення методу фотоемісійної спектроскопії з розділенням по куту (ARPES) суттєво збільшили кількість даних, що отримуються під час вимірювань. "Класичні" методи, такі як MDC- та EDC-аналіз або методи цифрової обробки зображень, не дозволяють швидко та ефективно обробляти значні обсяги отриманих даних. У статті проведено огляд існуючих методів обробки спектрів на основі згорткових нейронних мереж (ЗНМ), що дозволяють ефективно знешумлювати спектри та визначати електронну дисперсію. Крім того, запропоновано низку перспективних застосувань методів на основі ЗНМ для задач, що виникають при обробці фотоемісійних спектрів.

Ключові слова: фотоемісійна спектроскопія з розділенням по куту, згорткові нейронні мережі, машинне навчання.