IET Nanodielectrics

Special issue
Call for Papers

Be Seen. Be Cited.
Submit your work to a new
IET special issue

"High-performance
polyimide dielectric
materials"

Lead Guest Editor: Jun-Wei
Zha
Guest Editor: Lei Zhai

Read more

50% APC Discount code IET50
on submission.

IET The Institution of
Engineering and Technology

ORIGINAL RESEARCH

The Institution of Engineering and Technology WILEY

Identification of nanocomposites agglomerates in scanning electron microscopy images based on semantic segmentation

Yu Bai[1] ⓘ | Yan Wang[1] | Dayuan Qiang[2] | Xin Yuan[1] | Jiehui Wu[1] | Weilong Chen[1] | Sai Zhang[1] | Yanru Zhang[1] | George Chen[2]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

[2]The Tony Davies High Voltage Laboratory, University of Southampton, Southampton, UK

Correspondence

Yan Wang, School of Computer Science and Engineering, University of Electronic Science and Technology of China, No.2006 Xiyuan Ave West Hi-Tech Zone, Chengdu, China.
Email: yanbo1990@uestc.edu.cn

Abstract

The agglomeration causes a significant challenge to nanodielectrics. Identification of agglomerates in scanning electron microscopy (SEM) images is an important step of solving this issue. Motivated by the fast development of image recognition in computer vision, we propose a new approach for agglomerates identification in SEM images of nanodielectrics by semantic segmentation, which is more efficient and accurate than traditional methods. Three models based on convolutional neural networks are investigated in this work, namely pixel blocks classification network, full convolutional segmentation network employed with data augmentation and unsupervised self-encoding network. All three networks can preliminarily identify agglomerates of spherical silica-based blend polyethylene nanocomposites. The mean intersection over union (mIoU) of pixel blocks classification network is 0.843 and it takes 25 s to process an image. Full convolutional segmentation network only needs 0.059 s to process a sample, with a mIoU of 0.777. Unsupervised self-encoding network can reach a mIoU of 0.747 at a speed of 5.806 s. According to the amount of data sets, and requirements for different speed and accuracy, three kinds of networks can be flexibly selected.

KEYWORDS
dielectric materials, scanning electron microscopy

1 | INTRODUCTION

Polymer nanocomposites or nanodielectrics have been promised as future dielectrics [1, 2]. During the last decades, many experiment results exhibited that nanodielectrics can improve electrical properties in contrast to neat polymers [3]. For instance, 2 wt% MgO/low density polyethylene nanocomposites can enhance the Direct Current breakdown strength approaching 50% [4]; 0.5 wt% SiO_2/blend polyethylene nanocomposites can effectively decrease the surface potential decay [5] and suppress the accumulation of space charge [6]. Homogeneously dispersing nanoparticles into polymers is crucial to fully releasing the potential of nanodielectrics and improving electrical properties [2]. However, agglomeration as a major obstacle of industrialization scale application and further research on the interphase of

nanodielectrics can significantly affect the homogenous dispersion of nanoparticles. Figure 1 shows agglomerates (inside yellow contour lines) and matrix (dark parts) in a real scanning electron microscope (SEM) image of nanosilica-based polyethylene nanocomposites [7]. Agglomerates not only worsen the properties of nanodielectrics, such as moisture absorption and low breakdown strength [8], but also reduce the reliability of the experiment results [9]. To the best of our knowledge, there is no approach which can completely eliminate agglomerates in nanodielectrics, and approaches currently used can only mitigate agglomerates based on the qualitative observation. Therefore, observing and identifying agglomerates becomes one of the key tasks in the research of nanodielectrics, which is beneficial to correlate quantitative results of agglomeration to electrical properties of nanodielectrics.
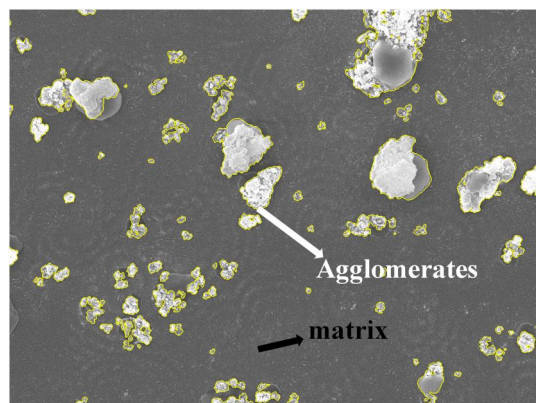
FIGURE 1 Agglomerates and matrix in a scanning electron microscopy image

Transmission electron microscopy (TEM) and SEM are commonly used for imaging nanocomposites. The effective observation range of TEM is less than 1 micron, while sizes of agglomerates published in recent years are generally between 2 and 10 microns [10–12]. A small observation scale may increase the randomness and subjectivity of sampling, which leads to higher uncertainty and inconsistency in final experimental results. SEM has an observation range of hundreds of picometres to hundreds of microns, which is more suitable for analysing the characteristics of agglomerates. By employing SEM images, a few methods have been proposed [11] to identify agglomerates, which first use binarization to identify agglomerates in SEM images and then quantify the degree of agglomerates. By setting a suitable threshold, binarization can straightforwardly distinguish agglomerates and polymers in images. However, this method cannot effectively deal with SEM images of semi-crystalline polymers-based nanocomposites. Firstly, the morphology of the crystalline region and agglomerates is similar after binarization, which dramatically increases the difficulties of identifying agglomerates by observing. Secondly, results of binarization can be easily affected by nonuniform brightness and contrast [11]. Therefore, the binarization cannot achieve expected identification of agglomerates in nanodielectrics.

In the realm of image recognition, the convolutional neural network (CNN) has shown excellent performance in recognizing objects from images [13]. In 1998, the LeNet-5 [14] was proposed as a highly feasible CNN architecture. With the development of computer hardware, such as continuous optimization of the graphics processing unit (GPU), more complex CNN frameworks can be realized. In 2012, AlexNet [15] also used the CNN architecture, which is similar to LeNet-5 but with more convolutional layers, and won the ImageNet large scale visual recognition challenge [16]. In 2015, ResNet [17] firstly obtained a higher accuracy of image recognition than humans in the competition. CNN has gradually been accepted by both industry and academia as its accuracy of image recognition was increasing [18]. Nowadays, CNN has become a widely applied high-performance method for a variety of image processing tasks. In order to meet the

requirements of more complex tasks, such as recognizing specific objects or distinguishing of pixels with different semantic information in images, more sophisticated network structures have also been proposed. In 2015, the Full Convolutional Network (FCN) [19] was proposed, which implements end-to-end semantic segmentation. FCN can classify all pixels in images at one time, greatly improving the running speed of semantic segmentation algorithms. Semantic segmentation has been applied into autopilot and medical image segmentation [18]. These algorithms can accurately process the pixel-level classification of pictures, which can save a lot of labour costs. Recognition of agglomerates in nanodielectrics is a similar task to the common image recognition, as shown in Figure 2. Hence, it also can derive a benefit from both CNN and FCN. However, according to our current literature research, the value of CNN and other deep learning algorithms has not yet received full attention in nanodielectrics research.

In this paper, three semantic segmentation methods are proposed to identify agglomerates in SEM images of spherical-silica based blend polyethylene nanocomposites. Considering the small amount of images we have collected, three networks are implemented, including pixel blocks classification network, full convolutional segmentation network employed with data augmentation and unsupervised self-encoding network. The main purpose of this work is to improve the identification of agglomerates in nanodielectrics SEM images by using the proposed methods instead of binarization.

2 | EXPERIMENTAL METHODS

A SEM image data set is built in this work, which can be used for experiments of deep learning algorithms. According to characteristics of SEM images, three kinds of network architectures are used to identify agglomerates.

2.1 | Preparation of specimen and data set

The spherical nanosilica was obtained from Sigma-Aldrich, and the range of its size is from 10 to 20 nm. Four loading ratios including 0.5 wt%, 2 wt%, 5 wt% and 10 wt% are applied in this investigation. The host polymer blend was 20% high-density polyethylene with 80% low density polyethylene. The specimens were etched before being imaged using SEM, and the etching procedure was performed based on a standard permanganic reagent [7]. After etching processing, all specimens were put on the aluminium stub and sputter-coated with gold. A Jeol JSM6500F high resolution FEG-SEM was used to measure all specimens at 15 kV.

A total of 28 SEM images with 1280 × 960 resolution are selected from the previous experiment results to construct an original data set. The SEM magnification range of the images in this data set is 2000–10,000. All images are annotated, that is, manually label agglomerates in each image and generate corresponding segmented images. Among all images, nine are used for training and the rest of them are used to test the
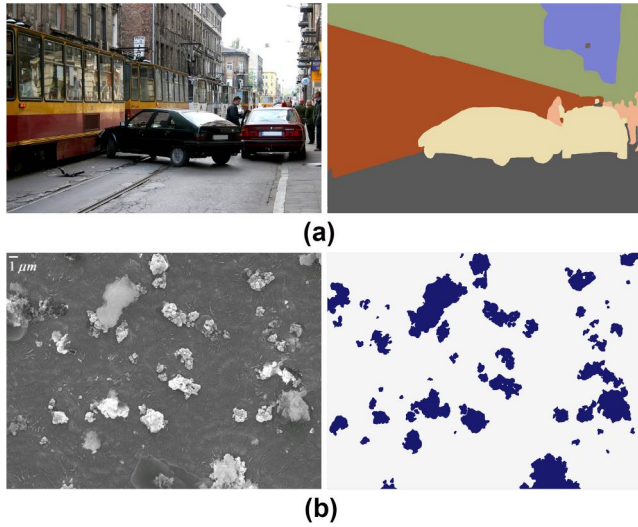
FIGURE 2 Semantic segmentation (a) A common image and its segmentation map, (b) A scanning electron microscopy image of nanosilica-based polyethylene nanocomposites [7] and its segmentation map

performance. Figure 3 shows specimens with 2 wt%, 5 wt%, and 10 wt% as examples. Images on the left column are original SEM data, and images on the right column are the corresponding annotated maps. In annotated maps, pixels in SEM images are labelled to two categories: matrix and agglomerates.

Based on this data set, designing corresponding semantic segmentation networks is the core task of this paper. The main obstacle is that SEM images are difficult to obtain in contrast to common images and the amount of data is small.

2.2 | Pixel blocks classification network

The pixel block classification network classifies every pixel in images one by one. Original SEM images are split into multiple pixel blocks, which are real classification objects. The number of pixel blocks is much larger than that of images, which can avoid the problem of fewer training samples.

2.2.1 | Pixel block

In this model, original SEM images need to be pre-processed. For each pixel in an image, we crop pixels in a certain range around it to generate a new small image, which is called a pixel block. The pixel in the centre of a pixel block is called a centre pixel. Figure 4 shows the cropping method of the pixel blocks. The label of the centre pixel is considered as same as the label of its corresponding pixel block. Therefore, we transform the image segmentation task into a classification task for all pixel blocks in images.

The cropping range of pixel block in this paper is 25×25, which is the resolution of the generated pixel block samples. 1,228,800 pixel blocks can be generated by a 1280×960 SEM
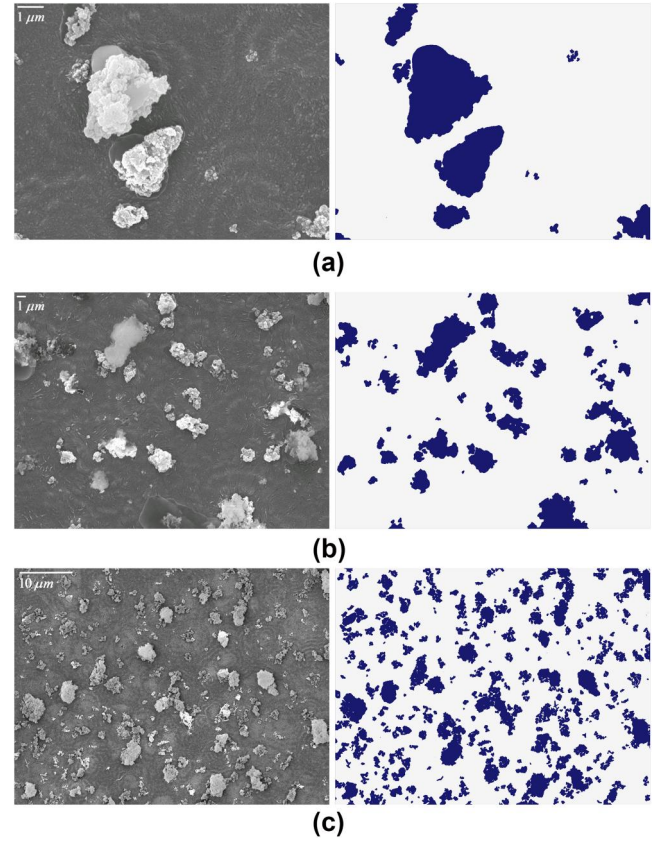

FIGURE 3 Scanning electron microscopy (SEM) images paired with annotated segmentation maps (a) A SEM image of 2wt% nanocomposites with magnification of 10,000, (b) A SEM image of 5wt% nanocomposites with magnification of 3500, (c) A SEM image of 10wt% nanocomposites with magnification of 2000
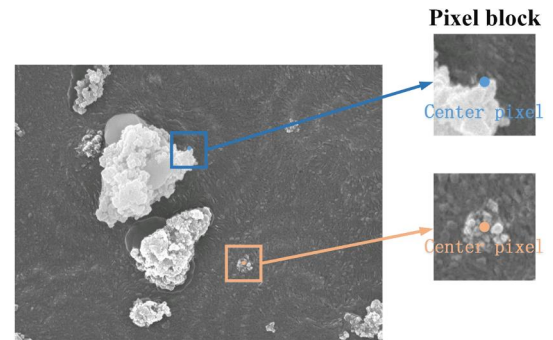

FIGURE 4 Centre pixel and pixel block

image. The size of the pixel block is a hyperparameter. We compared the experimental results of three different sizes of pixel blocks (15×15, 25×25, and 35×35), and selected the best size 25×25.

Images need to be padded to ensure that pixels at the edge of the images can generate its pixel block. Mirror padding is adopted in this paper, as shown in Figure 5. The black line is the boundary of the original image. The pixels inside the image boundary (the blue part) are mirrored and filled to the outside of the original image (the orange part). Compared with some
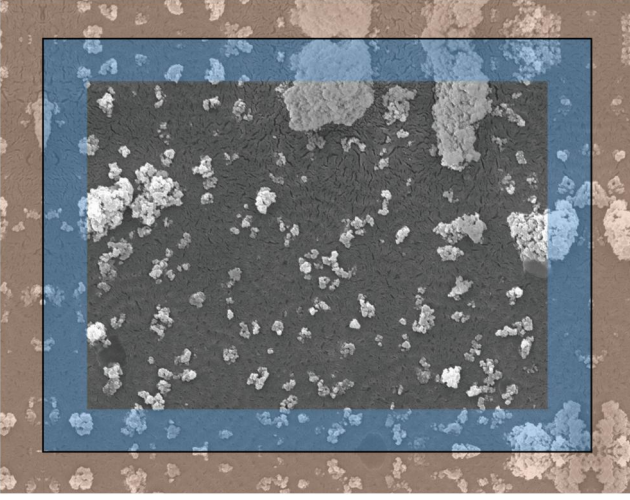
FIGURE 5 Mirror padding

common methods of directly padding 0 as pixel values, mirror padding can ensure that the pixel block corresponding to each central pixel has better information for training.

Next, a classification CNN network is designed to classify pixel blocks.

2.2.2 | CNN model

The designed CNN architecture is shown in Figure 6. Inputs of this network are pixel block, and outputs are their categories. Therefore, this network is called the pixel block classification network in this paper.

In this model, the network consists of three composite layers and three fully connected layers. The composite layer in this model contains two convolutional layers (convolution kernel size of 3×3, padding step size of 1) and one pooling layer (2×2 range maximum pooling) [14, 20]. The structure of the double convolutional layer cascade followed by maximum pooling is based on the construction method of VGGNet [21], which can better extract pixel block information features. When the size of the feature map becomes 3×3, the fully connected layer is used to further integrate features. After passing through two fully connected layers, pixel block features are mapped into two dimensions of agglomerates or matrix. The activation function of each layer is the rectified liner unit (ReLU) and the last layer uses softmax.

In order to illustrate the output of each layer, we give the following definitions. Outputs of the convolutional layer are k feature maps. Each feature map f is calculated by taking the dot product between the kth convolution kernel w^k of size $n \times n$, $w \in R^{n \times n \times k}$, and a local region x of size $m \times m$ with c number of channels, $x \in R^{m \times m \times c}$. The kth feature map $f \in R^{(m-n-1) \times (m-n-1)}$ is calculated as:

$$f_{ij}^k = \sigma \left(\sum_d^c \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} w_{abd}^k x_{i+a,j+b}^d \right) \tag{1}$$

where i, j represent the position of each pixel, and σ is the nonlinear activation function ReLU.

The pooling layer computes the maximum over a local non-overlapping spatial region in feature maps. The output $g \in R^{(m-n-1)/p \times (m-n-1)/p}$ of the pooling layer is calculated as:

$$g_{ij}^k = \max \left\{ f_{1+p(i-1),1+p(j-1)}^k, \cdots, f_{pi,1+p(j-1)}^k, \cdots, \right.$$
$$\left. f_{1+p(i-1),pj}^k, \cdots, f_{pi,pj}^k \right\} \tag{2}$$

where p is the size of the square pooled area and $1 \leq i, j \leq (m - n + 1)/p$.

After classification, the segmented image can be obtained by arranging all categories according their centre pixel positions one by one.

2.3 | Full convolutional segmentation network

The pixel block classification network requires pre-processing and need to process a large number of pixel blocks, which costs a lot of time. Inspiring by the FCN, a fully convolutional segmentation network is adopted in this paper. FCN is designed for realizing the end-to-end mapping, which can directly output a segmented map of an input image. The FCN model generally includes two processes: down-sampling and up-sampling. The down-sampling is to extract image features using backbone networks, which continuously shrink image feature maps due to operations such as convolution. The up-sampling is a process of restoring the resolution of feature maps, which is used to output the segmented image with the same resolution as the original image. This model is analysed from the following three parts.

2.3.1 | Backbone network

The backbone network adopted in this paper is ResNet50 [17], which can extract features of images effectively through its unique residual module. Only its convolutional layers are used in this paper.

2.3.2 | Pyramid pooling

It should be noted that sizes of agglomerates in SEM images vary greatly. Large-step pooling in the FCN model will integrate features in a large area of image, which can identify agglomerates with large size efficiently. However, many agglomerates with small scale may be ignored. To deal with this issue, the pyramid pooling module is adopted to identify multi-scale agglomerates in SEM images [22]. Figure 7 shows the structure of the pyramid pooling module.

This module performs multi-scale pooling at the tail of the backbone network to retain multiple scale features. Pooling
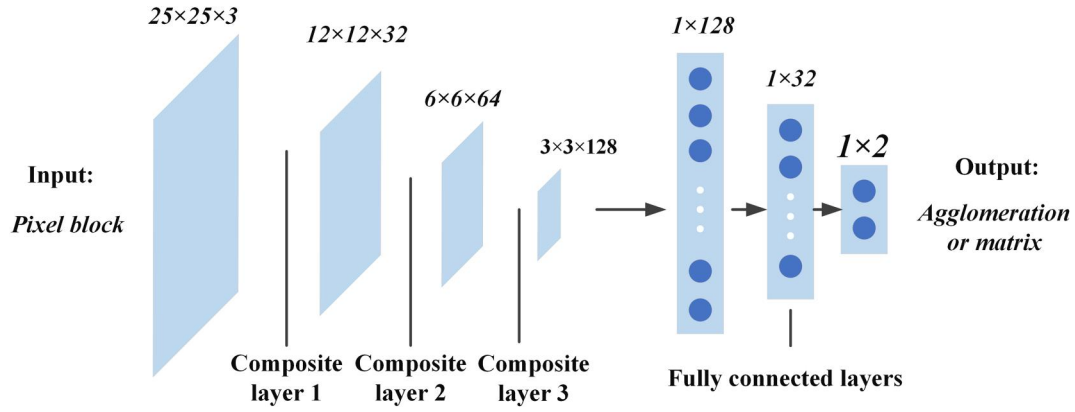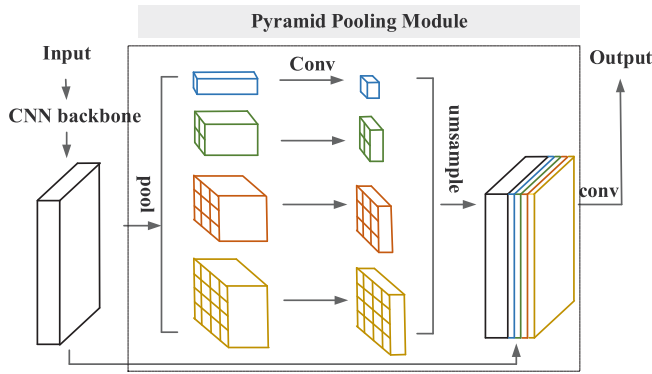
FIGURE 6 Pixel blocks classification network

FIGURE 7 Pyramid pooling module. CNN, convolutional neural network

steps includes 1×1, 2×2, 3×3, and 6×6. Four feature maps with different resolutions are obtained by multi-scale pooling. 1×1 convolution is used to reduce the number of channels of feature maps and refine features. Bilinear interpolation is used as the up-sampling method to expand the resolution of small-scale feature maps. Finally, we combine those feature maps with uniform size into a multi-channel feature map. The feature map undergoes subsequent convolution and up-sampling to obtain the final segmented image.

2.3.3 | Data augmentation

The full convolutional segmentation network has many parameters, which need a lot of SEM images for training. Data augmentation has been adopted in order to increase the number of SEM images and to avoid either underfitting or overfitting. Methods of data augmentation adopted are listed in Table 1.

The range of rotation in this paper is $(-25°, 25°)$. In order to keep the resolution of the rotated image same as the original image, we use pure black filling, which will not affect the agglomerates identification. Cropping is used to delete columns or rows of pixels on the side of the image. After cropping, the image resolution will be reduced. Interpolation

TABLE 1 Image augmentation

algorithms will be used to expand the resolution of the cropped image to the resolution of the original SEM image. This paper adopts random cropping in the range of 0%–20%, which means that the four sides of the image are cropped within the range of (0%–20%). Mirroring mainly includes horizontal and vertical mirroring. Rotation 90° means left or right 90-degree rotation without changing the resolution. After rotation, down-sampling is used for the long side, and bilinear interpolation is used for the short side to ensure that the image resolution remains unchanged. Due to the particularity of SEM images, no colour related augmentations have been applied.

2.4 | Unsupervised network

The two networks outlined above require annotated SEM image data sets for model training, which is called supervised learning in machine learning. However, it is difficult to obtain many SEM images practically, and manually annotating SEM images is also a time-consuming task. Hence, unsupervised learning network is adopted in this part, which does not rely on manual annotation of the SEM data set.

The segmentation process of the unsupervised model in this paper is divided into two parts. Firstly, a super pixel segmentation algorithm is used to pre-segment the input image. Secondly, a self-encoding network is used to merge the super pixels to get the segmented map.

Super pixel segmentation is a kind of image segmentation method based on clustering algorithms, which can divide an image into many regions. Pixels in these regions must have similarities in some aspects, such as colour or texture. These regions are known as super pixels. Super pixels usually do not destroy boundary lines of objects with semantic information in the original image. Simple linear iterative clustering [23] and Felzenszwalb [24] are two widely used super pixel segmentation algorithms. The pre-segmented results of the same SEM image through these two methods are shown in Figure 8. We need the pre-segmentation algorithm to generate enough super pixels with fine edge information. The result of Felzenszwalb has more complex boundaries as shown in Figure 8b, which is adopted in this paper.

After pre-segmentation, the self-encoding CNN is adopted to merge super pixels and get the segmented map [25, 26].

A self-encoding network is generally a fully convolutional network, which can make the resolution of the output image consistent with the input image. The self-encoding network can retain spatial features and optimize results of the pre-segmented map.

The architecture of the self-encoding network is shown in Figure 9, which refers to the structure of SENet [27]. The convolution kernels of 3×3 and 1×1 are alternately used in the network. The resolution of the feature maps keeps unchanged through padding, however, only the number of channels varies between 64 and 32.

The execution process of the unsupervised model is shown in Figure 10, which is divided into five steps:

Step 1: Pre-segmentation. The original SEM image (denoted as I) is pre-segmented by the Felzenszwalb algorithm. Then we obtained a pre-segmented SEM image with multiple super pixels, denoted as P.

Step 2: Extracting image features through the self-encoding network. A feature map with multiple channels is generated after the self-encoding network. Then, we only reserve the maximum value among all channels of each pixel position to get a segmented image with only one channel, which is recorded as F.

Step 3: Merging. We already have obtained two images, the pre-segmented image P and the self-encoded image F. For each super pixel in P, we count the category label m that appears most frequently in the corresponding area in F,
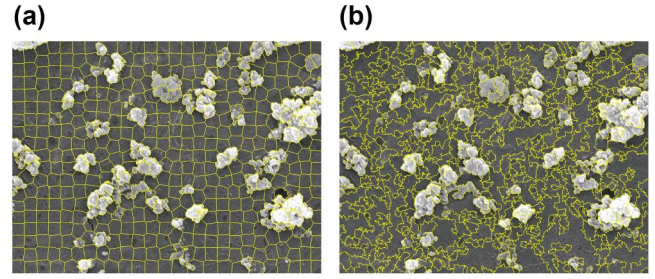

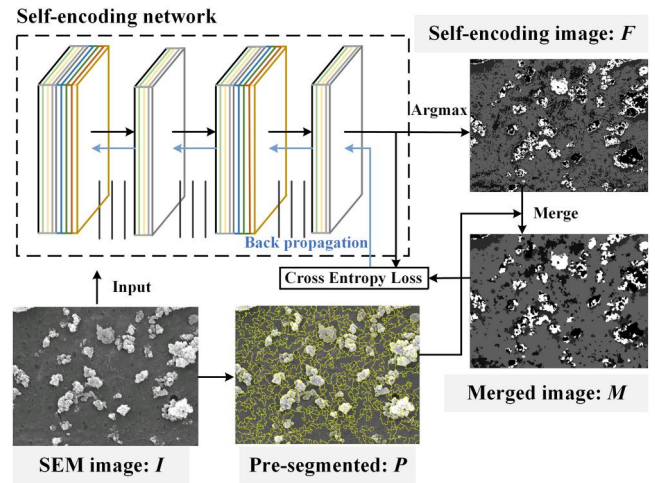**FIGURE 8** Results of two superpixel algorithms. (a) Simple linear iterative clustering, (b) Felzenszwalb

FIGURE 9 Unsupervised network. SEM, scanning electron microscopy

and then directly set all pixels in the super pixel range in F to m. After counting all super pixel regions and replacing the pixel values one by one, we obtain the merged image denoted as M.

Step 4: Iterating and updating the network. The merged image M is used as the label of the original SEM image to train the self-encoding network. The loss function adopted is cross entropy. Then the model performs back propagation and updates its parameters. After that, step 2 is performed again to obtain a new self-encoded image F_1, and then step 3 and 4 are performed using F_1. This model optimizes the network by iteratively executing steps 2, 3, and 4.

Step 5: Generating a two-class segmentation image. After n iterations, the final self-encoded image F_n is obtained. In F_n, the SEM image is segmented into an image with multiple categories of regions. The clustering algorithm cannot accurately segment all pixels in the image to a specific number of categories, especially the extreme two-category. According to pixel areas of different categories in F_n, we calculate the average value of each area pixel in the original image I, and the pixel values of the area are replaced with the average value to obtain the image FI. FI is not a two-class segmentation image, but can be compared to the original image to confirm agglomerates and matrix. For automatic two-class classification,
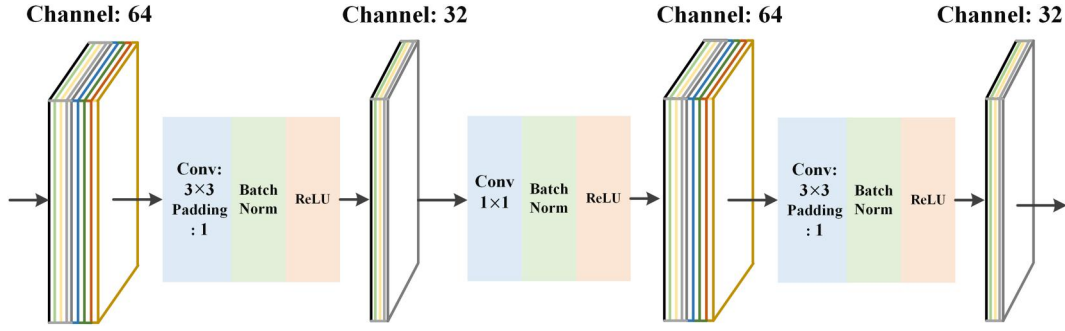
Channel: 64   **Channel: 32**   **Channel: 64**   **Channel: 32**

F I G U R E 1 0 The structure of the self-encoding network

we count the average pixel value in *FI* and use it as a dichotomous threshold to divide agglomerates and base material to get the final two-class image *S*.

In the unsupervised network, the super pixel segmentation algorithm provides fine-grained region division. The self-encoding network merges super pixels to generate a better segmented map.

3 | RESULTS AND DISCUSSIONS

Experimental results of three networks are presented in this section. The performance of models is analysed from three aspects, including the accuracy of agglomerates identification, loss curve and time consumption. In addition, this paper compares the effect of semantic segmentation model and traditional binarization method on SEM images of semi-crystalline polymers-based nanocomposites. And the effect of magnification on the recognition accuracy is tested. At the end, the applicability of our models is discussed.

3.1 | Identification accuracy of the agglomerates

The pixel accuracy (PA) and the intersection over union (IoU) are commonly used metrics in semantic segmentation [13]. In order to illustrate these two metrics, this paper gives the following definitions. The number of pixel categories is $k + 1$ ($0 - k$, including background categories). p_{ij} is the number of type i pixels that are misjudged as the type j. p_{ii} is the number of type i pixels that are correctly judged as type i.

3.1.1 | Mean pixel accuracy (MPA)

The PA is the ratio of the number of correctly classified pixels to the total number of pixels. MPA is the average value of PA of all categories.

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \qquad (3)$$

T A B L E 2 MIoU and MPA of the three methods

Methods	MIoU	MPA
Pixel blocks classification network	0.843	0.917
Full convolutional segmentation network	0.777	0.871
Unsupervised network	0.747	0.844

Abbreviations: MIoU, mean intersection over union; MPA, mean pixel accuracy.

3.1.2 | Mean intersection over union

IoU is an indicator for calculating the degree of overlap between two sets. Mean IoU (MIoU) is the average value of IoU of all categories. The value of MIoU ranges from 0 to 1.

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \qquad (4)$$

All the three networks can preliminarily identify agglomerates of spherical silica-based blend polyethylene nanocomposites under the condition of a small number of SEM images. The MIoUs of pixel blocks classification networks, full convolutional segmentation network, and unsupervised network reach 0.843, 0.777, and 0.747, respectively, according to Table 2.

For exhibiting the performance of three networks intuitively, an example is depicted in Figure 11. The output segmented map of pixel blocks classification network shows more details similar to the label than the other two networks, which is in consistent with the results of MIoU.

3.2 | Loss curve

In order to verify the convergence of models, loss curves of the pixel block network and the fully convolutional segmentation network are shown in Figures 12 and 13. Figures 12a,b are the training and validation curves of the pixel block network. Figures 13a,b are curves of the fully convolutional segmentation network. Both networks have
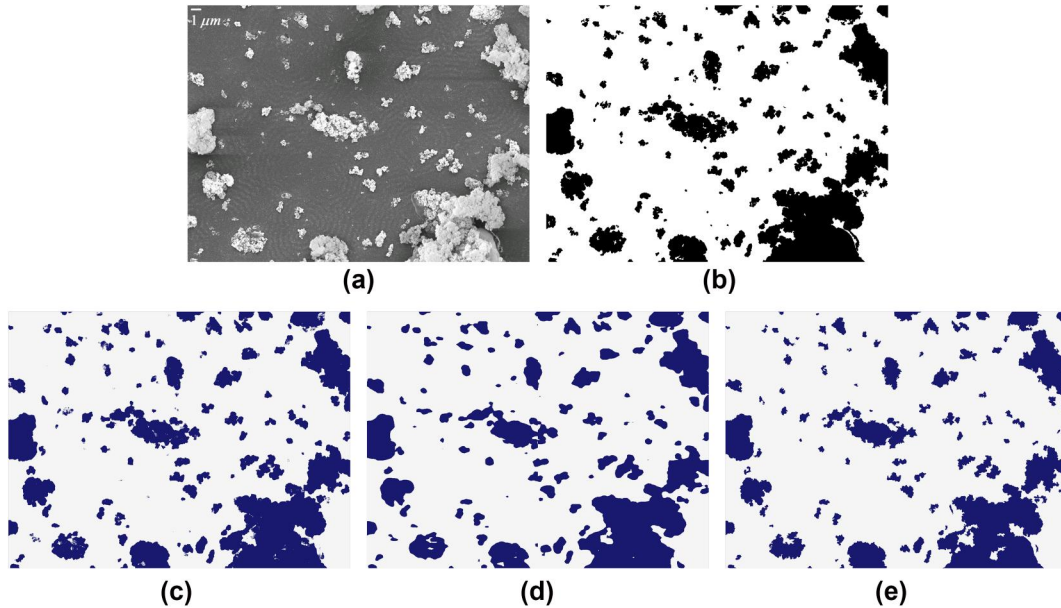
FIGURE 11 The performance of three networks. (a) Original image, (b) Label, (c) The result of the pixel network, (d) The result of the Full Convolutional Network, (e) The result of the unsupervised net

good convergence and are not overfitted. The unsupervised network does not need to be trained, so it has no similar curve.

3.3 | Time cost

The time cost of models is a valuable evaluation index. In actual semantic segmentation applications, in order to meet the requirements of the equipment, the time spent on model inference must be precisely controlled. It has an important analytical value for us to improve networks.

The environment of experiments in this paper is a sever with one GPU (NVIDIA GeForce RTX 2080 Ti, video memory: 11G) and CPUs (6×Xeon E5-2678 v3, memory: 62 G). The uniform size of the input SEM images is 1280 × 960. Table 3 lists the average time it takes for three kinds of networks to predict a SEM image.

The pixel block classification network takes much longer than the other two networks. The main reason is that the time cost of the pixel block network is directly proportional to the resolution of input SEM images. The pixel classification network needs to process all pixel blocks corresponding to pixels, which means that 1,228,800 pixel blocks need to be predicted for an image of 1280 × 960.

It is worth noting that, for the pixel block classification network, the batch size of inputs has a great impact on its performance. The resolution of a pixel block is much smaller than an entire image. Pixels blocks only occupy a quite low GPU memory, which means we can set the batch size in a very wide range. As long as the video memory is not exceeded, as the batch size increases, the number of pixel blocks processed by the network at the same time increases, which means that the processing speed of the

pixel block network is accelerated. At the same time, different batch sizes also have a certain impact on the prediction accuracy. Therefore, we can't increase the batch size limitlessly. In order to balance speed and accuracy, the batch size is 128 in this paper.

However, fully convolutional segmentation network can directly realize end-to-end semantic segmentation, which will not be affected by the resolution of input images. Hence, its running speed is extremely fast.

Although the self-encoding network in the unsupervised network is also a fully convolutional network, this model includes the process of super pixel segmentation and model self-training, which costs more than the fully convolutional segmentation network.

3.4 | Results on SEM images with crystalline regions

The inability to effectively process SEM images with crystalline regions is a major drawback of binarization methods. The binarization method may misjudge the crystalline texture as agglomerates, resulting in segmented images with larger errors.

The results of the binarization method and three networks we used are compared in this test. Figure 14 shows the results of different methods. Figures 14a,b are the original SEM image and its label. Figure 14c is the result of the binarization method with a threshold of 127. The crystalline area contains many black dots after binarization because the binarization only uses pixel values as the standard for judgement, ignoring the morphology of the crystalline region, light and dark changes and other factors. Three networks used in this paper can accurately determine the crystalline region, as shown in Figures 14d–f.
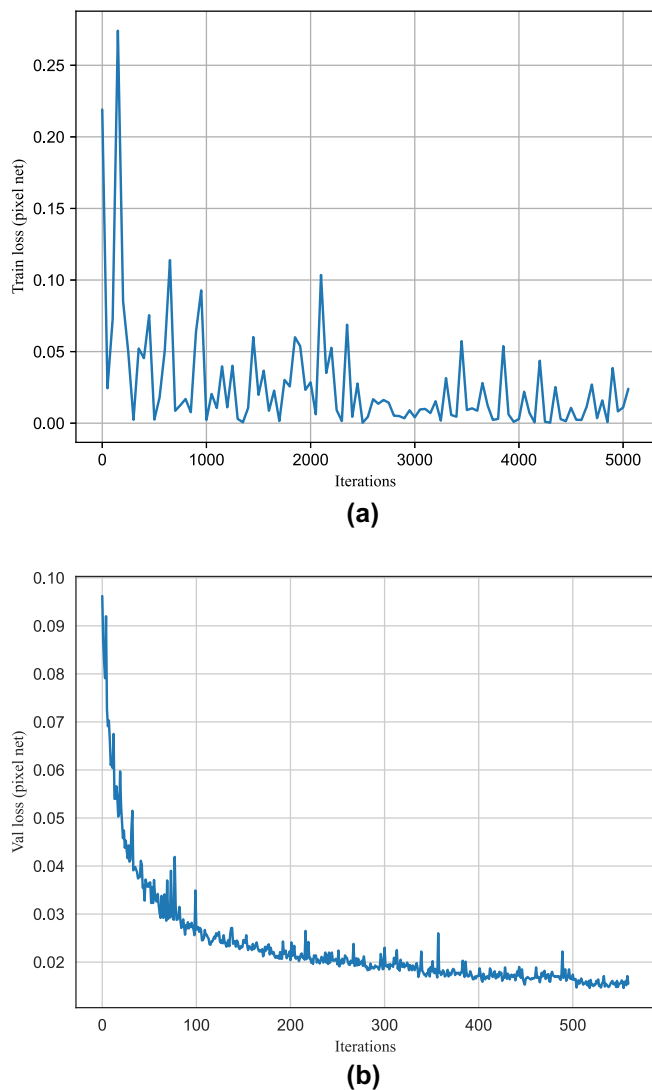
(a)

(b)

FIGURE 12 Loss curves of the pixel block classification network. (a) Train loss curve of the pixel block classification network, (b) Validation loss curve of the pixel block classification network

(a)

(b)

FIGURE 13 Loss curves of the full convolutional segmentation network. (a) Train loss of the full convolutional segmentation network, (b) Validation loss curve of the full convolutional segmentation network

3.5 | The effect of magnification on identification

Magnification is an important factor in analysing SEM images. The prediction accuracy of SEM images with different magnifications can reflect the generalization ability of the networks to a certain extent.

For studying this effect, two networks are trained with images of an observation magnification, and SEM images with different observation magnification are used for validation. The results of different magnification are depicted in Figure 15. It displays the results of the pixel block network. The other two networks can also work at multiple magnifications, and the results are similar to the pixel block network, which are omitted here. The magnification of the training images uses 3500 paired with the manually annotated segmentation map as the label, which is shown in Figure 15b. The tested images paired with the segmentation results are shown
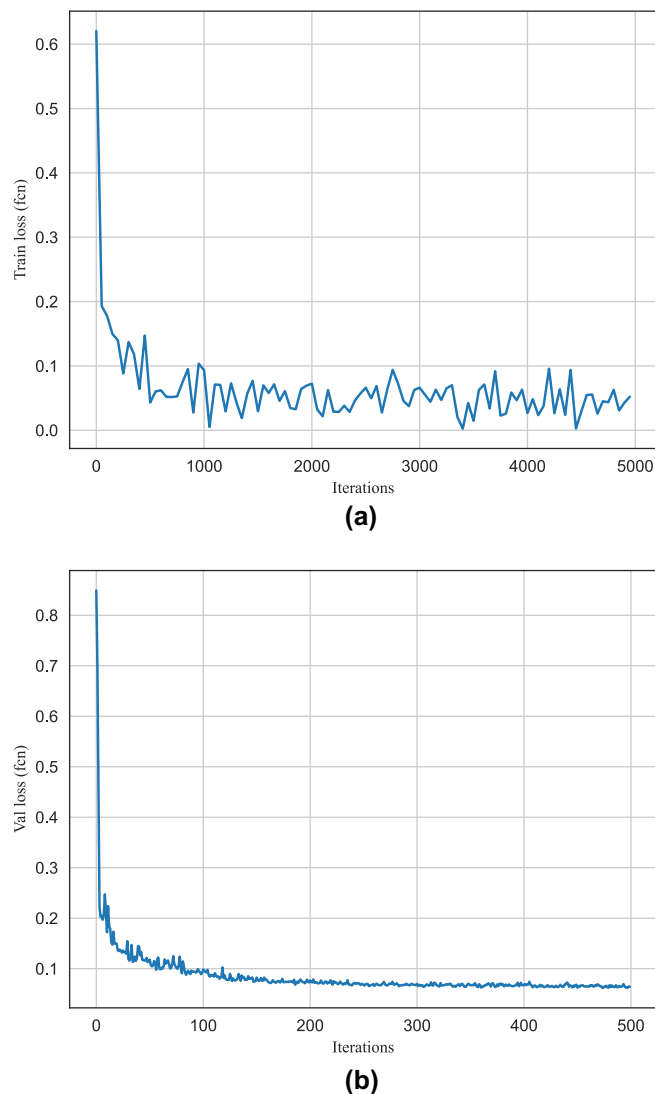
TABLE 3 Time cost of the three methods

Methods	Average time cost (s)
Pixel blocks classification network	25.00
Full convolutional segmentation network	0.059
Unsupervised network	5.806

in Figures 15a,c, of which the magnification is 10,000 and 2000, respectively. In addition, the image shown in Figure 15a is a part of Figure 15b which is also a part of Figure 15c as highlighted. The mIoU of 10,000 magnification and 2000 magnification can reach 0.906 and 0.863, respectively, which reflects that these networks can generally achieve multi-scale recognition of the spherical silica-based blend polyethylene nanocomposites. However, it is thought that training the networks with images containing more features of agglomerates such as images of 2000 magnification can obtain higher mIoU
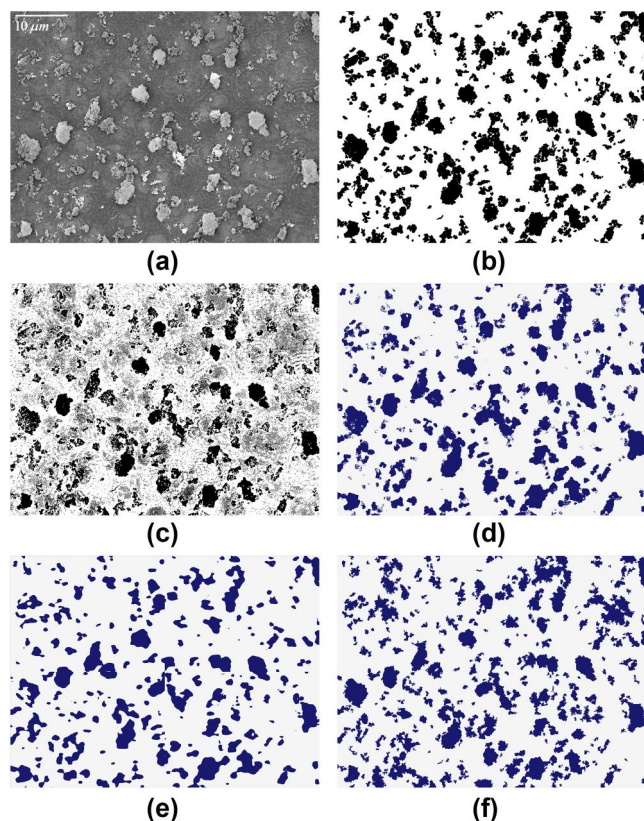
FIGURE 14 Results of a scanning electron microscopy image with crystalline regions. (a) Original image, (b) Label, (c) The result of binarization, (d) The result of the pixel network, (e) The result of the Full Convolutional Network, (f) The result of the unsupervised net

FIGURE 15 The effect of magnification on the identification of agglomerates. (a) Test image with magnification of 10,000 and segmentation result, (b) Train image with magnification of 3500 and label, (c) Test image with magnification of 2000 and segmentation result

when testing the images containing less features of agglomerates such as images of 10,000 magnification.

3.6 | Discussion and extension

This paper aims to identify agglomerates in SEM images. Considering the small number of SEM images, the pixel blocks classification network and the unsupervised model are adopted, which do not rely on a large amount of training data. Taking into account the speed and the wider applicability of the model, a fully convolutional network is used, whose performance is believed to be improved with more training data. Three semantic segmentation models output segmented images with only two categories of pixels: matrix and agglomerates. Single particles are not considered in this paper. In our data set, the magnification of SEM images ranges from 2000–10,000, which allows agglomerates be clearly observed. Single particles are almost impossible to distinguish individually in these images.

4 | CONCLUSION

In this work, we proposed a novel approach for identifying agglomerates of nanodielectrics by usingthe semantic segmentation algorithm. Under the condition of a small number
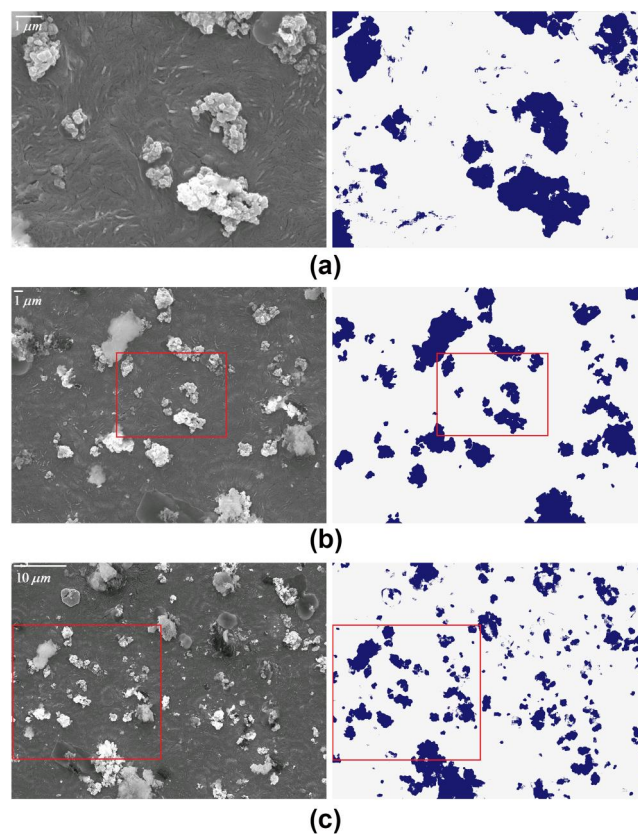
of SEM images, pixel blocks classification network, full convolutional segmentation network, and unsupervised network preliminarily realize the identification of agglomerates in SEM images of the spherical silica-based blend polyethylene nanocomposites. The first network has a slightly higher MIoU; the second network has dramatically fastened the running speed; and the third network does not depend on a data set for training. It is highly likely that the accuracy of the full convolutional segmentation network can be further increased by feeding more SEM images. In the future, more correlated SEM images will be collected for fully releasing the potential of the full convolutional segmentation network.

ORCID
Yu Bai https://orcid.org/0000-0002-3948-9224

REFERENCES
1. Lewis, T.: Nanometric dielectrics. IEEE Trans. Dielectr. Electr. Insul. 1(5), 812–825 (1994)
2. Zhong, S.L., et al.: Past and future on nanodielectrics. IET Nanodielectr. 1(1), 41–47 (2018)

3. Nelson, J.K., et al.: Dielectric polymer nanocomposites. Springer (2010)
4. Murakami, Y., et al.: Dc conduction and electrical breakdown of mgo/ldpe nanocomposite. IEEE Trans. Dielectr. Electr. Insul. 15(1), 33–39 (2008)
5. Wang, Y., et al.: Experimental demonstration of deep traps in silica-based polyethylene nanocomposites by combined isothermal surface potential decay and pulsed electro-acoustic measurements. Appl. Phys. Lett. 113(2), 022904 (2018)
6. Wang, Y., et al.: The effect of loading ratios and electric field on charge dynamics in silica-based polyethylene nanocomposites. J. Phys. Appl. Phys. 51(39), 395302 (2018)
7. Wang, Y., et al.: Dc current in nanosilica-based polyethylene nanocomposites. In: 2015 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP). IEEE, pp. 515–518 (2015)
8. Wang, Y., et al.: Influence of moisture absorption on electrical properties and charge dynamics of polyethylene silica-based nanocomposites. J. Phys. Appl. Phys. 51(42), 425302 (2018)
9. Wang, Y., Chen, G., Vaughan, A.: Space charge dynamics in silica-based polyethylene nanocomposites. In: 2014 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP). IEEE, pp. 727–730 (2014)
10. Ayoob, R., Andritsch, T., Vaughan, A.: The effect of material processing on the dielectric properties of polystyrene boron nitride nanocomposites. In: 2015 IEEE Electrical Insulation Conference (EIC). IEEE, pp. 333–336 (2015)
11. Del Gaudio, C., Licciardi, G.A.: A simple tool for two-dimensional quantification of filler dispersion: a proposal. Fullerenes, Nanotub. Carbon Nanostruct. 27(5), 446–452 (2019)
12. Guo, M., et al.: Polyethylene-based composites containing octaisobutyl polyhedral oligomeric silsesquioxanes obtained by extrusion. In: 2015 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP). IEEE, pp. 527–530 (2015)
13. Garcia, A., et al.: A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:170406857 (2017)
14. LeCun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE. 86(11), 2278–2324 (1998)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105 (2012)
16. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115(3), 211–252 (2015)
17. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
18. Khan, A., et al.: A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53(8), 5455–5516 (2020)
19. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
20. Längkvist, M., et al.: Classification and segmentation of satellite orthoimagery using convolutional neural networks. Rem. Sens. 8(4), 329 (2016)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 (2014)
22. Zhao, H., et al.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
23. Achanta, R., et al.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34(11), 2274–2282 (2012)
24. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. Int. J. Comput. Vis. 59(2), 167–181 (2004)
25. Kanezaki, A.: Unsupervised image segmentation by backpropagation. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp. 1543–1547 (2018)
26. Kim, W., Kanezaki, A., Tanaka, M.: Unsupervised learning of image segmentation based on differentiable feature clustering. IEEE Trans. Image Process. 29, 8055–8068 (2020)
27. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)