International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp.1-14

https://doi.org/10.70917/ijcisim-2026-0014
Article

Research on Image Segmentation Algorithm Based on Sparse Representation and Multi-Task Learning

Aiwu Chen *

College of Intelligent Manufacturing, Hunan University of Science and Engineering, Yongzhou 425199, Hunan, China; caiwu9050@126.com

Abstract: The results of image segmentation greatly affect the accuracy and correctness of image recognition. In this paper, image segmentation is taken as the research focus, combing and introducing the principles of multi-task learning, image sparse representation and tracking algorithm. An image segmentation model based on sparse representation and multi-task learning is designed, and an attention learning network AL-Net based on multi-task learning is proposed. The loss function of multi-task makes the contour of the segmented image smoother. The image segmentation model constructed in this paper is applied to the image segmentation of cucumber disease leaves, when the number of images in the training sample rises from 80 to 120, the recognition accuracy of the model in this paper stably stays at 90%, and the average recognition rate of seven diseases, such as leaf spot and scab, is as high as 90.49%, and the average time consumed is only 7.68s, which makes the effectiveness remarkable.

Keywords: multi-task learning; image sparse representation; tracking algorithm; image segmentation

1. Introduction

Images provide an effective way to analyze information about a cognitive scene readily, quickly, efficiently, and from multiple perspectives, which not only provide intuitive visual digital representations of specific objects and overall overviews of a scene, but also describe geometric properties such as the location, shape, and structure of objects in a given scene [1-3]. These properties are often hidden in the spectral energy distribution within the image theta, and they are the basic units of cognitive scene content [4-5]. No matter how complex the scene represented by an image is, the brain can accurately and quickly recognize which regions embody the content and which objects carry the key information, and analyze and understand the image content based on the objects and their spatial contextual relationships [6-7]. Image segmentation is one of the image techniques, which segments different regions of an image according to their special meanings, and there is no correlation between different regions, and all specific regions have consistency [8].

Researchers simulate the visual function of the brain proposed a series of image segmentation algorithms, and widely used in medical analysis, aerospace, unmanned aerial vehicles, remote sensing observation, augmented reality, etc., to assist the completion of the task in various fields, and therefore in the continuous breakthrough in the performance of segmentation technology [9-11]. For example, in the medical field, doctors often utilize CT, MR and other medical images to assist in diagnosing the condition and improve the accuracy of diagnosis and the effectiveness of treatment plans [12-13]. In the field of aerospace, the use of relevant segmentation technology to process and analyze remote sensing images collected by artificial satellites and aircraft, to complete resource surveys, disaster monitoring, resource surveys, agricultural planning, and urban planning and other operations, have achieved good results [14-17]. In the field of unmanned driving, road planning, obstacle avoidance and other functions can be accomplished by correlating and analyzing the images obtained from various image acquisition devices such as cameras [18-19].

Some common image segmentation methods are as follows: literature [20] in order to deal with high

noise and blurred edges in magnetic resonance imaging of osteosarcoma, the Transformer and U-net model were introduced to design a professional image segmentation method, and the method was further optimized by adding an edge enhancement module and a loss function. Literature [21] found that medical image segmentation with weighted fuzzy kernel clustering algorithm outperforms the fuzzy kernel clustering algorithm to minimize the classification error and noise. Thresholding is a commonly used image segmentation technique, but its segmentation quality has been the focus of research. Literature [22] proposed a new meta-heuristic balancing algorithm to optimize the thresholding of grayscale images and outperforms whale optimization algorithm, particle swarm optimization algorithm, etc. in terms of segmentation accuracy within the task threshold. Literature [23] utilized texture features of MRI images, combined their first and second order statistical feature vectors for sparse coding and developed dictionary learning kernel clustering algorithms to classify them and obtained higher quality of image segmentation.

And U-net is a commonly used model for image segmentation algorithms, but its encoder and decoder organization architecture has poor cross-scale sparse correlation in image segmentation, high computational resource requirements, and limited scope of applicability [24]. Therefore scholars have gone ahead and proposed a number of methods to optimize this model. Literature [25] developed a new U-net architecture, GA-UNet, using genetic algorithm for medical image segmentation of lungs, livers, etc., with an average accuracy of 98.58%, friendly to a large amount of computational resources. However, there is no literature to make optimization and new methods proposed for its sparse association. Literature [26] segmented multi-task fundus images with end-to-end deep neural network, which is realized by learning the middle layer and the last layer of segmentation from the original image and is based on U-net architecture. Literature [27] gives an adaptive algorithm for optimal path snakes for high and low resolution medical image segmentation, which somewhat outperforms some of the algorithms and has wide applicability.

In addition, literature [28] segmented real images based on manual rice RGB color image pixel classification using full convolutional network and its semantic segmentation method and U-Net model, and the accuracy of the full convolutional network semantic segmentation method was the best. Literature [29] has optimized full convolutional network to enhance the image texture by acquiring local and frequency domain features of biomedical sensing images, which are further enhanced by combining the comparison of weights of target regions. However, the full convolutional network has significant detail texture loss for unstructured local information processing. And literature [30] designed a multi-task quadruple attention network to optimize the phenomenon of local region misclassification in remote sensing image segmentation, which improved the image segmentation accuracy by 3.57%-6.33% with quadruple attention (position, label, channel, and edge), combined with multi-task mechanism.

However, these researches still have not addressed the problems of cross-scale sparse association and the synergy between semantic segmentation and instance segmentation in multitasking. And sparse representation is a technique that can utilize the properties of the signal itself to describe the signal. For a K-dimensional signal z, if it can be expressed in the form of a combination of N-dimensional basis vectors, then it is a sparse signal, and the use of sparse representation techniques can greatly reduce the redundancy of information in image reconstruction, thus improving the efficiency of image processing [31-32]. Multi-task learning improves the generalization ability and performance of the model by learning multiple related tasks simultaneously. Compared with single-task learning, multi-task learning can better utilize the limited labeled data and improve the model's adaptability to different tasks by sharing representations and parameters [33]. Literature [34] trained the tongue image as a dictionary with local blocks of the training image and sparse representation with the blocks of its dictionary, obtained the tongue probability through sparse, formed a probability map, and optimized the edge information of the tongue image of similar color, which improved the performance of tongue image segmentation. As for image segmentation with multi-task learning, literature [35] takes chest X-ray images of lung nodules as an example, showing that the quality of image segmentation is improved mainly by learning the images of lung regions and labeling the data in the region, which contributes to the accuracy of lung nodule recognition. Based on this, sparse representation and multi-task learning will become important techniques for the optimization of image segmentation algorithms.

In this paper, the principles of multi-task learning, image sparse representation and tracking algorithms are introduced respectively, which provides a theoretical basis for the construction of image segmentation model, and constructs an image segmentation model based on sparse representation and multi-task learning. An image segmentation network (AL-Net) with multi-task attention learning is proposed to improve the feature extraction ability of the main task. A context encoding layer is added behind each layer of the encoder, and its output is processed by the attentional learning module and fused with the features of the decoding layer to reduce over-segmentation and under-segmentation. Reduce over-segmentation and under-segmentation. The shape loss function is utilized to speed up the

convergence of the contours after image segmentation and to make them smoother. Compare and analyze the image segmentation quality of the image segmentation model in this paper on common image segmentation public datasets to explore the model performance. Carry out cucumber disease leaf image segmentation application practice to further analyze the effectiveness of the image segmentation model of this paper in realistic image segmentation tasks.

2. Image Sparse Representation and Multi-Task Learning Algorithm

The study of image segmentation algorithms based on sparse representation and multi-task learning, this chapter will introduce the principles of multi-task learning, image sparse representation and tracking algorithms involved [36].

2.1. Principle of Image Sparse Representation

The sparse representation algorithm based on redundant dictionary can portray the detailed information of the image well, and can represent the main features of the image with a small number of coefficients, i.e., the sparse representation is to express the original image by utilizing a relatively small number of non-zero elements in the dictionary.

Let the original image be represented by a column vector $y \in R^n$ and matrix $D = \{d_1, d_2, \cdots, d_m\} (D \in R^{n \times m}, n < m)$, then y can be expressed as:

$$y = D\alpha \tag{1}$$

where D is the overcomplete dictionary and each column vector is called an atom of the dictionary. $\alpha \in R^m$ is the sparse representation coefficients of the image y on the dictionary D, containing few nonzero elements.

Solving this sparse representation model can be viewed as solving its L_0 -paradigm problem, i.e.,:

$$\min \| \alpha \|_0 \ \ s.t. \ \| y - D\alpha \| \leq \varepsilon \tag{2}$$

In the above equation, ε is the sparse approximation error; $\| \alpha \|_0$ is the L_0 -paradigm of α, which indicates the number of non-zero elements. Due to the non-convexity of the L_0 -paradigm, it is a typical NP-hard problem to solve Eq. (2) exactly, so it can be converted to the L_1 -paradigm to find its approximate solution. That is, problem (2) is transformed into solving problem (3):

$$\min \| \alpha \|_1 \ \ s.t. \ \| y - D\alpha \| \leq \varepsilon \tag{3}$$

In solving this problem, researchers often use tracking algorithms to find the approximate solution of the sparse coefficients, and some of the more commonly used sparse optimization decomposition algorithms are matching tracking, orthogonal matching tracking, and basis tracking. The most commonly used algorithms in recent years are matching tracking and orthogonal matching tracking.

2.2. Tracking Algorithm

2.2.1. Matching Tracking Algorithm

The matched tracing algorithm (MP algorithm for short) is an algorithm for sparse decomposition of signals over an overcomplete dictionary, which is one of the more popular sparse decomposition algorithms in recent years [37].

Assume that H is a Hilbert space, the source signal is $y \in H$, the overcomplete dictionary is $D = \{x_1, x_2, \cdots, x_{n_0}, \cdots, x_i\} \{i \in (1, 2, \cdots, k)\}$, and the atom has been made normalized, i.e. $\| x_i \| = 1$. The steps of the algorithm are roughly as follows:

Step1: Initialize the residuals. Let $R_0 f = y$, where $R_0 f$ refers to the initial residual.

Step2: Select the most matching atom. Perform the inner product operation between the source signal and each atom in the dictionary, and select the atom of which makes the absolute value of the operation result the largest. The selected atom needs to satisfy the following equation:

$$|\langle y, x_{r_0} \rangle| = \sup_{i \in (1,2,\cdots,k)} |\langle y, x_i \rangle| \tag{4}$$

where x_{r_0} refers to the r_0 th atom in the dictionary.

Step3: Iterate to find the residuals. Use the following equation to find the new residual $R_1 f$, i.e., the signal Y is decomposed into two parts: the orthogonal projection component on x_{r_0} and the residual value:

$$R_1 f = R_0 f - \langle y, x_{r_0} \rangle x_{r_0} \tag{5}$$

The $\langle y, x_{r_0} \rangle x_{r_0}$ is the orthogonal projection of Y on the atom x_{r_0}, and $R_1 f$ is the residual value.

Step4: Perform Steps 2 and 3 for the residual values, which can be obtained after K iterations:

$$R_k f = \langle R_k f, x_{r_{k+1}} \rangle x_{r_{k+1}} + R_{k+1} f \tag{6}$$

where $x_{r_{k+1}}$ is satisfied:

$$|\langle R_k f, x_{r_{k+1}} \rangle| = \sup_{i \in (1,2,\cdots,k)} |\langle R_k f, x_i \rangle| \tag{7}$$

At this point, the signal Y is decomposed into:

$$y = \sum_{n=0}^{k} \langle R_n f, x_{r_n} \rangle R_n f + R_{k+1} f \tag{8}$$

where $R_0 f = y$.

This algorithm is able to obtain an approximate solution of the source signal, but it is computationally intensive, and because the re-projection operation in the subspace that has been selected to consist of atoms is non-orthogonal, the result of each iteration is not necessarily optimal, and it is prone to over-matching after many iterations, so an orthogonal matching tracking algorithm has been introduced.

2.2.2. Orthogonal Matching Tracking Algorithm

The Orthogonal Matching Tracking algorithm (called OMP algorithm) is a greedy algorithm that works iteratively by greedily selecting support atoms in order to recover the sparse signal. At each step, the highest absolute inner product between the support atoms and the residuals is found and the corresponding atoms are added to the already found support atoms, and then the new residuals are computed.

The improvement of the OMP algorithm over the MP algorithm is that all the selected atoms are Schmidt orthogonalized each time, which ensures that the result is optimal after each iteration, and therefore improves the efficiency of the algorithm.

Suppose the source signal is $y \in H$ and the overcomplete dictionary is $D = \{x_1, x_2, \cdots, x_{n_0}, \cdots, x_i\} \{i \in (1, 2, \cdots, k)\}$. The steps of the algorithm are roughly as follows:

Step1: Initialize the residual $e_0 = y$.

Step2: Follow the atom selection principle of the MP algorithm so that the selected atom x_0 satisfies equation (4).

Step3: Form the selected atoms into a column matrix X_t, and define the orthogonal projection operator in the column space of X_t as:

$$P = X_t (X_t^T X_t)^{-1} X_t^T \tag{9}$$

Step4: Calculate the residual e_1:

$$e_1 = e_0 - Pe_0 = (I - P)e_0 \tag{10}$$

$$e_{m+1} = e_m - Pe_m = (I - P)e_m \tag{11}$$

Step5: Perform steps 2~4 iteratively on the residuals, while iterating it is important to note that atoms that have already been selected will not be selected again.

Step6: Until the specified stopping criterion is reached, the algorithm stops. This algorithm can get the optimal result in each iteration compared to the MP algorithm, which reduces the number of iterations in this paper, the OMP algorithm is used in experiments to solve the sparse coefficients.

2.3. Principles of Multi-Task Learning Algorithm

Most machine learning tasks to date have been single-task learning. However, many problems in real life cannot be subdivided into individual subproblems, and even if they can be subdivided, the correlation information between each subproblem will be lost. Therefore, in order to obtain better learning results, it is necessary to decompose the complex system into several simple subsystems before solving, which is one of the main research contents of multi-task learning theory. But the traditional multi-objective optimization algorithm is difficult to achieve the acquisition of the optimal solution. Multi-task learning is a problem that arises when multiple tasks share the weights of some characteristics that can be shared in the learning process, which has a better generalization effect than single-task learning.

2.3.1. Multi-Task Learning Structure

Multi-task learning is a machine learning method based on shared representation to learn multiple related tasks simultaneously [38]. Multi-task learning is the derivation method of migration learning, which is to do continuous derivation bias of the information related to the learning region in the training signals of related tasks in order to improve the generalization of the main task. It maps the relationship between different tasks to the same model for computation and prediction to improve the performance of the operating system. The proposal of multitasking can be considered as a major breakthrough to the traditional machine learning approach and will be widely used in various fields. The learning process of multitasking learning involves multiple tasks such as learning multiple corresponding tasks in parallel, back-propagating gradient information, and using low-level shared representations to help each other, learn and enhance generalization effect capabilities. In short, the multitasking learning process refers to learning multiple corresponding tasks at the same time (note: they must be related tasks, and the definitions of the related tasks and the information they share need to be given). The learning process then refers to the simultaneous sharing of each other's information from a shallow common representation and filling in each other's learning scope to improve learning efficiency and thus generalization.

2.3.2. Parameter Sharing Mechanisms

Multi-task learning shared representations aim to improve generalization, the simplest shared method for multi-task learning, where multiple tasks share parameters on a shallower layer. There is correlation between different tasks, which requires a separate shared representation for each subtask. This approach not only ensures high accuracy but also maintains high efficiency.There are two ways of MTL sharing:

1)Parameter sharing (based on parameters): e.g., neural network based MTL, Gaussian process.

2) Constraint sharing (rule-based): e.g., mean value, fusion feature learning (setting up a common feature pool).

Most multitask learning research today can be categorized into two groups: one is multitask learning network design: the other is multitask learning loss function design. Network design for multitask learning can be categorized into two cases: hard sharing mechanism of parameters and soft sharing mechanism of parameters.

Making shared parameters is one of the basic baselines of almost all multitask learning, which encodes the task information in the whole backbone network as a shared network. At the last layer the network is split into several task-specific decoders to do the prediction. Making the sharing mechanism a lower bound on the number of parameters in the network design is an important reference for balancing efficiency and accuracy in designing the network and reduces the risk of overfitting.

The soft sharing mechanism for parameters can be considered as the other extreme of the hard sharing mechanism for parameters and is not commonly compared today in the present multitasking network design. In the soft sharing mechanism for parameters, each task has a parameter space and a backbone

network. The similarity between sparse gradients is compared by imposing specific constraints on their parameter spaces. In this paper, a density matrix decomposition based method is utilized to estimate the correlation coefficients and residual variance array between each unknown function in the model; these coefficients are then solved by an iterative algorithm. The distance between model parameters is regularized to ensure the similarity of model parameters.

The design of the loss function for multi-task learning is designed to go for better updating of task-specific gradients in the network.

For multi-task learning of task $i(i=1,2,...,N)$ for multi-task learning, the loss function is defined as:

$$L=\sum_{i}^{N}\lambda_i L_i \tag{12}$$

where λ_i are task-specific learning parameters. It is necessary to find a good set of λ_i to optimize the L_i parameters for all tasks i.

3. Image Segmentation Model Based on Sparse Representation and Multi-Task Learning

The traditional image segmentation model has shortcomings such as high computational cost and low segmentation efficiency. To address this problem, this paper will design an image segmentation model based on sparse representation and multi-task learning in order to realize the improvement of image segmentation accuracy and segmentation efficiency.

3.1. Structure of AL-Net

Based on the multi-task model structure of parameter sharing + each task has its own independent attention module (MTAN), the shared parameter part is a complete predictive model connecting each task's attention module for each layer. The output of a particular layer of a task is the weighted average of the output of that layer in the shared network and the attention of the current task with the corresponding element.MTAN improves the generalization ability of the shared parameters by adding the attention to select the shared parameters while making the parameters shared.

This chapter proposes an attention learning network based on a hard parameter sharing multitasking framework. In particular, this chapter designs a novel attention learning mechanism that enhances the learning of attentional weights.The network structure of AL-Net is shown in Fig. 1.Firstly, an image is input into an encoder which includes a pre-trained ResNet-34 and a context encoding layer. The context encoding layer is added after each pooling in order to extract contextual features at different resolutions. The obtained features are then fed into an attention learning module. The results of this module are fused with high-dimensional features as jump connections, and the shape loss for the main task is computed after decoding is complete. The outputs of all the attention modules are upsampled to the size of the input image in order to compute the loss for the secondary task. Finally, the weighted sum of the losses of the two tasks is the total loss of the network.

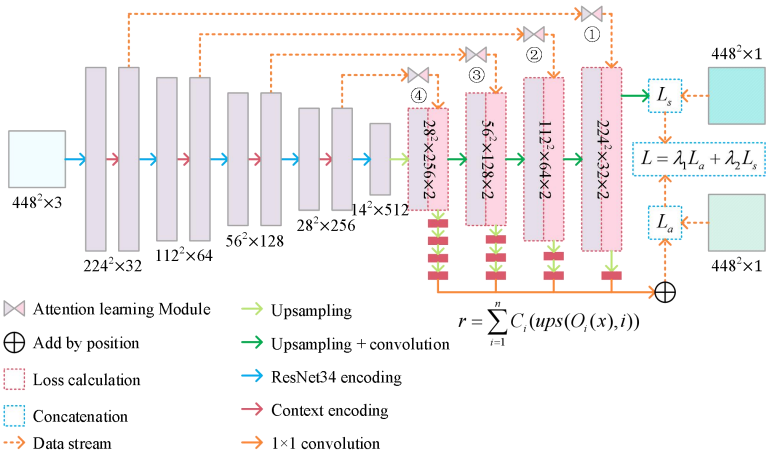

Figure 1. Network structure and loss function of AL-Net.

3.1.1. Context Encoding Module

In image segmentation, feature extraction includes convolution and pooling operations. Pooling can increase the sensory field of the convolution kernel and improve the computational efficiency. However, some spatial information is always lost in the process of downsampling and then upsampling. Null convolution solves this problem to a certain extent. Null convolution is the process of expanding the convolution kernel by adding some spaces or zeros between the elements in the convolution kernel, which expands the receptive field without adding extra parameters.

The Dense Cavity Convolution (DAC) module of CE-Net obtains spatial information at different scales through four cascaded multiscale cavity convolution branches. The Residual Multiscale Pooling (RMP) module introduces the idea of pyramid pooling, pools different sizes of cavity convolution, and is able to predict the spatial information of different resolutions, which solves the problem of losing spatial information by successive convolution operations. And keep the number of parameters unchanged while increasing the receptive field to capture more global semantic information, and the cascade branching widens the range of convolutional operations, the large receptive field convolution is better at detecting large target objects and extracting features with a higher degree of abstraction, and the small receptive field convolution is more advantageous in the detection of small target objects, and the combination of the two kinds of convolutions can be applied to sufficiently improve the network segmentation accuracy.

In this chapter, a context coding layer is placed after each coding layer in order to extract context features with different resolutions.

3.1.2. Attention Learning Mechanisms

Suppose that it is necessary to compute a query vector q for a set of inputs $H=[h_1,h_2,h_3,...,h_n]$, obtained after passing through the attention mechanism, when a query vector q related to the task is needed, and a scoring function is used to compute the relationship between the query vector q and each input h_i to obtain a score. These scores are then normalized using the softmax function, and the normalized result is the distribution of the attention of the query vector q over each input h_i $a=[a_1,a_2,a_3,...,a_n]$, where each value is associated with the original input $H=[h_1,h_2,h_3,...,h_n]$ correspond one to one. Taking a_i as an example, the relevant formula is as follows:

$$a_i = softmax(s(h_i,q)) = \frac{\exp(s(h_i,q))}{\sum_{j=1}^{n} exp(s(h_j,q))} \tag{13}$$

Suppose again that the input information is no longer in the form of $H=[h_1,h_2,h_3,...,h_n]$, but rather in the more general form of key-value pairs, $(K,V)=[(k_1,v_1),(k_2,v_2),...,(k_n,v_n)]$, the associated query vector remains q. In this mode, the query vector q and the corresponding key k_i are generally used to compute the attention weight a_i:

$$a_i = softmax(s(k_i,q)) = \frac{\exp(s(k_i,q))}{\sum_{j=1}^{n} exp(s(k_j,q))} \tag{14}$$

The attention mechanism emphasizes or extracts important features and suppresses irrelevant details. The input feature map is first upsampled and then a single channel feature map is obtained using a codec. The feature map is then normalized using Sigmoid function in order to obtain the attention weights. Finally residual join method is used to obtain the output feature map $O(x)$ with the following formula:

$$O(x) = F(x) \oplus (F(x) \otimes A(x)) \tag{15}$$

where: $F(x)$ denotes the input feature mapping (value term), $A(x)$ denotes the attentional weight, $\oplus$ denotes the summation by elements, and $\otimes$ denotes the product by elements. In the attention learning module, the upsampling module enlarges the boundary of the overlapping kernel to reduce the loss of spatial information. The encoder extracts the salient features of the image and the decoder learns the

attention weights from the model training.

3.2. Loss Function for Multitasking

The designed network consists of two branches. The loss function L is defined as:

$$L = \lambda_1 L_a + \lambda_2 L_s \tag{16}$$

where: L_a is the loss function for the auxiliary task and L_s is the shape loss for the primary task. λ_1, λ_2 denote the weights of the loss function. L_a is defined as follows:

$$L_a = L_{bce}(r, b) \tag{17}$$

where: L_{bce} denotes the binary cross-entropy loss function. r denotes the fusion result of all attention modules and b denotes the labeled contour image. r is defined as:

$$r = \sum_{i=1}^{n} C_i(ups(O_i(x), i)) \tag{18}$$

where: $C(x)$ denotes a 1 × 1 convolution, $ups(x, i)$ denotes that x is upsampled i times, and $O_i(x)$ denotes the output feature map of the ith attention module.

The principle of the shape loss function is to add a penalty term to the image region in order to aggravate the loss in the error region. The formula is:

$$L_s = L_{bce}(c, g) + P \tag{19}$$

where: P is the penalty term, c denotes the output of the network and g denotes the truth value. By minimizing this weight, the prediction result is close to the true value. Take an image as an example to illustrate the computation of the penalty term, the region surrounded by the green curve in the figure is the prediction result, the yellow region is the true value, and the edges of the prediction result are divided into n regions by square boxes ($n = p / w$, p is the number of contour points, and w is the size of the square boxes). First, the weight I_i of each box is calculated as:

$$I_i = \frac{A \cap B}{A \cup B} \tag{20}$$

where: A denotes the area of the true value region in the square box and B denotes the area of the predicted region in the square box. The closer the value I_i is to 1, the closer the predicted area is to the true value. Then, the shape feature F (ellipticity) of the target is calculated. We want F to be close to 1. The penalty term P is calculated as:

$$P_i = (1 - I_i)(1 - F) \tag{21}$$

where: the penalty value of the predicted region in the ith square box is set to P_i. The penalty value of the other pixels is set to 0. In this way, the prior of the target shape is added to the loss function. This function can effectively control the shape of the contour and smooth the contour.

3.3. Experiments and Analysis of Results

3.3.1. Experimental Environment and Parameter Configuration

Hardware configuration of the experiment: 16G video memory, 64G RAM, Intel core i7 CPU, NVDIA GeForce GTX1060 GPU.

Experimental software configuration: Windows 10 operating system, using Python 3.6 programming language and Pytorch deep learning framework.

3.3.2. Experimental Data Sets

This experiment uses common public datasets for image segmentation: the MS COCO dataset and the Cityscapes dataset.

1) MS COCO dataset

The MS COCO dataset is characterized by rich contextual information in the images, and some information in the images can be obtained from the object categories, numbers, and instances of the MS COCO dataset, which contains a variety of different categories of objects. Compared with other datasets, the total number of object categories is less, but each category contains more instances of different shapes and sizes of the objects.

2) Cityscapes dataset

Cityscapes dataset is a relatively small-scale dataset, the images mainly show city street scenes collected from 50 different cities in different seasons, so it is also called cityscape dataset.

3.3.3. Evaluation Indicators

Image Quality, abbreviated as PQ.PQ can be split into two parts. One part is the Recognition Quality (RQ for short), which is used to represent the system's ability to recognize different instances. The other part is the segmentation quality (SQ for short), which is used to represent the system's ability to give correct masks to the predicted objects. In order to evaluate the segmentation ability of semantic branch and instance branch more clearly, this paper also uses PQ^{Th} and PQ^{St} to distinguish the performance of the two branches. Where PQ^{Th} denotes the system's ability to segment Thing-like objects and PQ^{St} denotes the system's ability to segment Stuff-like objects.

3.3.4. Analysis of Experimental Results

JSISNet, OANet, AUNet, HoVer-Net are chosen as the comparison models, and the results of this paper's model and other models in COCO dataset are compared as shown in Table 1. It can be seen that the accuracy of this paper's model is greatly improved compared with other models, with PQ reaching about 60%, and SQ and RQ being the highest 82.1% and 70.3%, respectively. In terms of the ability to segment Thing-type objects and Stuff-type objects, the model in this paper outperforms the other comparative models.

Table 1. Comparison of results of different models in COCO data set.

Model	PQ (%)	SQ (%)	RQ (%)	PQ^{Th} (%)	PQ^{St} (%)
JSISNet	26.5	71.4	36.6	30.3	24.4
OANet	41.2	77.6	50.5	49.8	27.1
AUNet	46.1	81.5	56.2	55.7	33
HoVer-Net	53.4	80.1	69.1	58.3	43.1
Model of this article	60.1	82.1	70.3	64.3	55

The comparison of the results of this paper's model with other models on Cityscapes dataset is shown in Table 2. It can be seen that the PQ of this paper's model is higher than that of other comparative models, which can reach 64.5%, and the SQ and RQ are 83.8% and 77.7%, respectively, which are also better than other comparative models. The performance of this paper's model is still excellent in the task of segmenting Thing-type objects and Stuff-type objects.

Table 2. Comparison of results of different models in Cityscapes data set.

Model	PQ (%)	SQ (%)	RQ (%)	PQth (%)	PQst (%)
JSISNet	46.9	75.4	59.2	39.9	50.5
OANet	59	50.3	47.2	54.6	61.6
AUNet	61.4	81.8	75.8	58.4	64.6
HoVer-Net	61.4	81.4	74.6	56.2	62.2
Model of this article	64.5	83.8	77.7	60.7	67.8

The loss function convergence curves of this paper's model on COCO dataset and Cityscapes dataset are specifically shown in Fig. 2. The convergence degree of this model on COCO dataset and Cityscapes dataset is better. Comparing the two curves, it can be seen that the convergence degree of this method is relatively better on the Cityscapes dataset, and the convergence value can reach about 0.1. The convergence value on the COCO dataset is about 0.2.

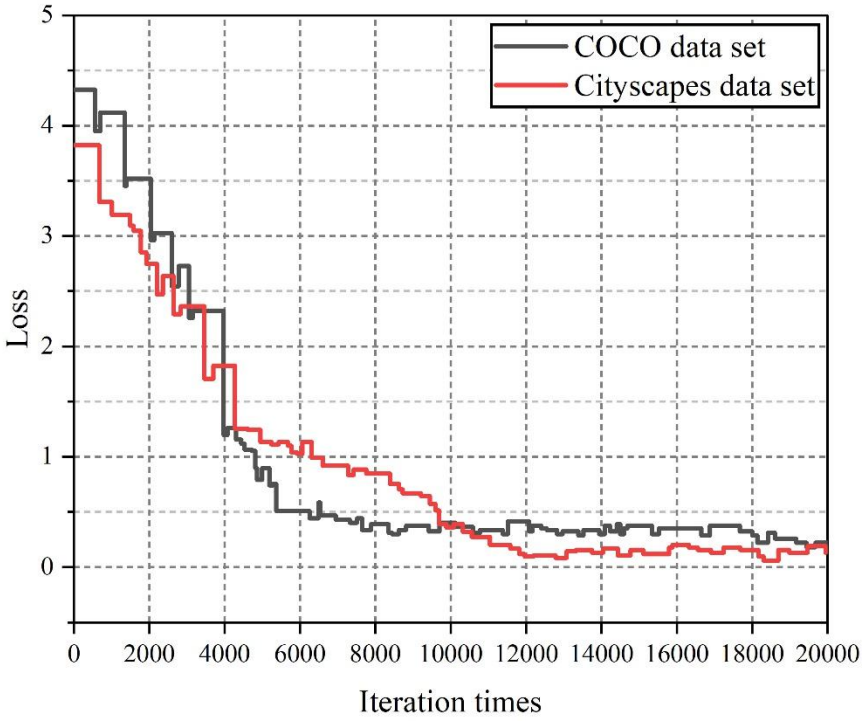

Figure 2. Convergence curve of loss function.

4. Cucumber Disease Leaf Image Segmentation Application Practice

In order to verify the effectiveness of the image segmentation model based on sparse representation and multi-task learning constructed in this work, and its practical utility in the real image segmentation work, in this chapter, we will select the cucumber disease leaf image to carry out the application practice, and compare it with the results of SVM, ANN, KNN, and SRC published in recent years.

The test platform for this application practice is Lenovo computer (Intel I7 processor, 2GB RAM, Windows2K, 64bit operating system).

4.1. Data set Construction

Images of cucumber leaves with seven diseases, namely downy mildew, angular mottle, leaf spot, scab, gray mold, anthracnose and powdery mildew, were taken at the farm attached to Zhongkai Agricultural Engineering College using a digital camera and a smartphone. The same shooting environment was set for all images in order to avoid interference from localized strong lighting and atmospheric turbulence. The image size was set to 256 × 256 pixels by cropping and scaling. All images were RGB three-channel color images and saved in BMP format. Two hundred images of each disease were obtained, and a total of 1400 samples were acquired. For each disease image, 120 images were randomly selected for model training, and the remaining 80 images were used for model performance testing. Considering the noise or uneven illumination in the shooting process, the images were smoothed and filtered and histogram equalized before the disease region segmentation, and then the disease images were feature extracted and the disease species labels were added to construct the training data set.

4.2. Analysis of Practical Results

The results of different models to recognize the test samples are specifically shown in Figure 3. In the case of considering only a single form of features, the model recognition algorithm in this paper is degraded into a sparse representation-based recognition algorithm, the sparse coefficient solution is similar to the SRC algorithm, and the recognition efficiency is lower than the SRC algorithm, which utilizes the theory of sparse representation classification for disease recognition, but this method stacks the color features, shape features, and texture features indistinguishably. The model algorithm in this paper comprehensively considers the nature of different form features, has stronger constraints on sparse coefficients, and the recognition results are more accurate and robust. Meanwhile, it can be found that the recognition efficiency of this paper's model also increases when the number of training samples increases, and it is always higher than other methods. In addition, when the number of each disease image in the

training samples rises from 80 to 120, the recognition rate of this paper's model is relatively stable at about 90%, and it can be considered that this paper's model relies less on the training samples, and under the condition of a smaller number of samples, it will still have a higher recognition efficiency.

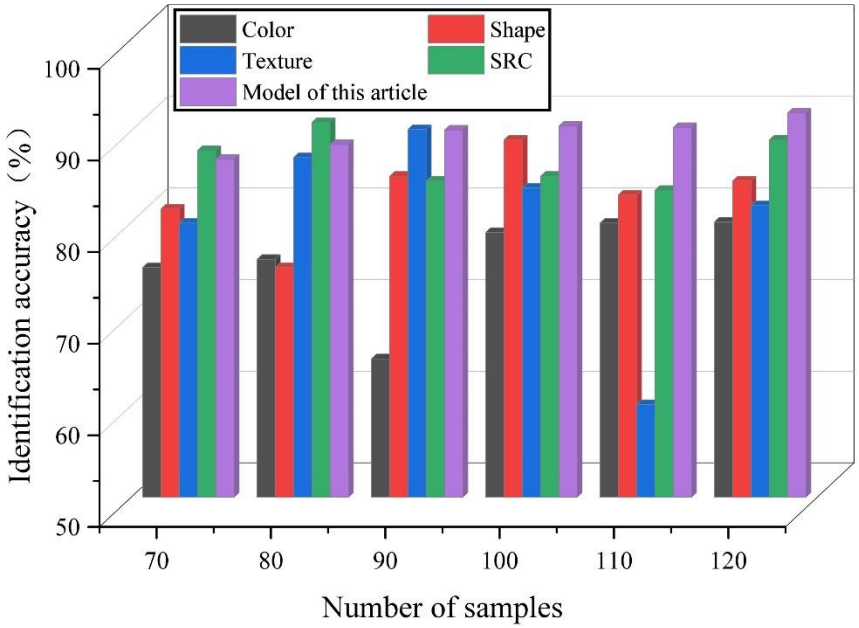

Figure 3. Accuracy rate of cucumber disease identification.

The number of training samples is set as 840 (i.e., each disease contains 120 training samples) for the experiment, and the comparison of the correct recognition rate and average consumption time of the diseases with the selected frontier algorithms SVM, ANN, KNN and SRC is shown in Table 3. It can be found that the model in this paper has the highest average recognition accuracy of 90.49% for these seven diseases, while the average consumption time is the least, only 7.68s.

Table 3. The recognition accuracy of joint sparse model algorithm.

Disease name	Identification accuracy (%)					Consumption time (s)				
	SVM	ANN	KNN	SRC	Model of this article	SVM	ANN	KNN	SRC	Model of this article
Downy mildew	81.44	81.83	82.4	88.93	89.31	20.4	8.64	35.62	10.57	7.75
Target spot disease	70.28	71.32	79.47	82.17	88.64	25.54	11.84	36.86	13.24	8.23
Leaf spot disease	72.31	72.55	77.26	84.42	90.17	23.72	11.55	37.48	13.53	7.92
Scab disease	73.35	70.48	73.12	83.3	87.98	22.62	12.58	38.29	13.34	8.22
Gray mold disease	85.24	87.74	86.16	91.23	95.11	20.97	8.66	35.38	9.94	6.49
Anthracnose	78.39	73.79	81.64	84.59	92.45	21.2	9.88	37.19	11.28	8.02
Powdery mildew	77.47	78.08	83.39	86.19	89.75	23.16	10.82	36.36	10.52	7.15
Average value	76.93	76.54	80.49	85.83	90.49	22.52	10.57	36.74	11.77	7.68

SVM, ANN, KNN and SRC algorithms all use a combination of color, shape, and texture features for disease recognition, and all of them simply stack the three types of features and calculate them as a unified feature vector, without considering the correlation relationship between different features. The sparse coefficients of arbitrarily extracted test samples under the model representation of this paper are specifically shown in Fig. 4. It can be seen that the sparse coefficients of the color, shape and texture

features of this paper's model are approximated to 0 in the vast majority of disease samples, and coefficients with larger values exist at the corresponding positions of a small number of samples and are concentrated in a few disease categories. More importantly, the sparse coefficients corresponding to the color, shape and texture features of the samples have a similar structure, which is the key to the ability of this paper's model to improve the efficiency of disease recognition.

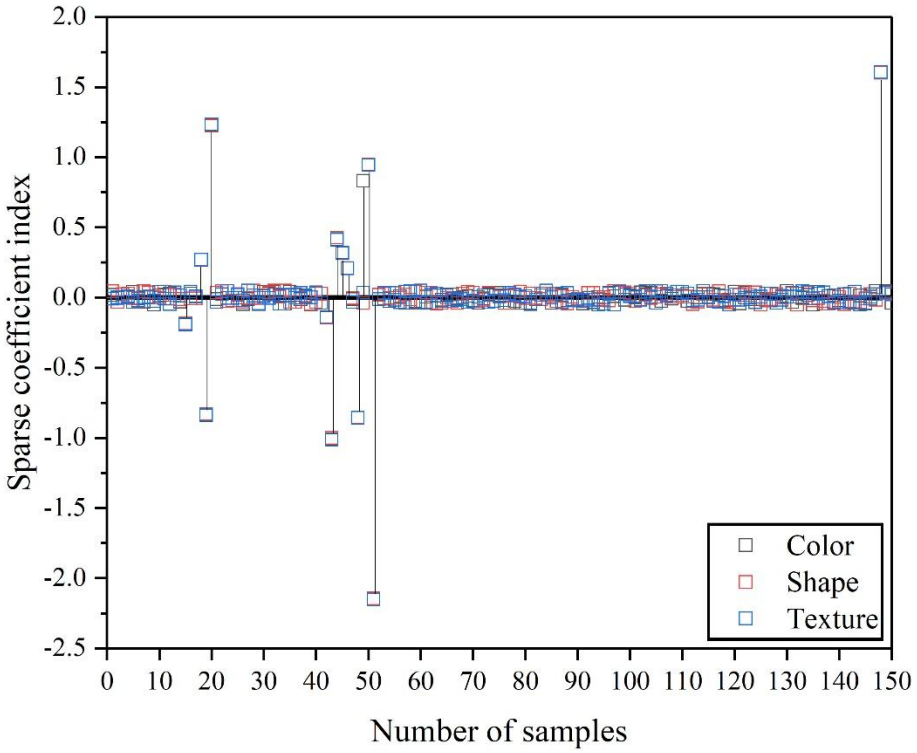

Figure 4. Joint sparse coefficient.

5. Conclusion

Aiming at the problems of high computational cost and low segmentation efficiency faced by traditional image segmentation methods, this paper establishes an image segmentation model based on sparse representation and multi-task learning to improve the segmentation performance. Model performance comparison experiments are carried out on MS COCO dataset and Cityscapes dataset, comparing with JSISNet, OANet, AUNet, HoVer-Net, the image segmentation model in this paper achieves a Q of about 60%, and the SQ and RQ are the highest among all the models, which are 82.1% and 70.3%, respectively, and the ability of segmenting the objects of Thing class is also better than other comparisons, Stuff class objects is also better than other comparison models. The convergence of the model in this paper is better on both COCO dataset and Cityscapes dataset.

The model is applied to the cucumber disease leaf image segmentation work to explore the practical application of the model. When the number of images of each disease in the training samples increases from 80 to 120, the recognition accuracy of the model in this paper can always be maintained at about 90%, which is still a high recognition efficiency under the condition of a small number of samples. Comparing with SVM, ANN, KNN and SRC in terms of the correct recognition rate and average consumption time of the disease, the model in this paper has the highest average recognition accuracy of 90.49%, and also has the shortest average consumption time of 7.68 s. The model in this paper has the highest average recognition accuracy of 90.49%, and also has the shortest average consumption time of 7.68s. Sampling any test sample, the sparse coefficients under the model representation of this paper are approximated to be 0 in the vast majority of disease samples, and only in a small number of samples corresponding to the position of the coefficients of larger values exist.

Overall, the image segmentation performance of the image segmentation model based on sparse representation and multi-task learning constructed in this paper is exc ellent, and it has good application effect in real image segmentation and recognition work.

Funding
The research is supported by the Research Foundation of the Natural Science Foundation of Hunan Province, (Grant No. 2024JJ7189); The Social Science Project of Hunan Provincial Achievement Review Association (Grant No. XSP24YBC319); Hunan Province General Higher Education Teaching Reform Research Project (HNJG-20231101); Hunan Province General Higher Education Teaching Reform Research Project (HNJG-20231094).

References

1. Singh, V., Girish, D., & Ralescu, A. L. (2017). Image Understanding-a Brief Review of Scene Classification and Recognition. MAICS, 2017, 85-91.
2. Cheng, G., Han, J., & Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE, 105(10), 1865-1883.
3. Ali, N., Zafar, B., Riaz, F., Hanif Dar, S., Iqbal Ratyal, N., Bashir Bajwa, K., ... & Sajid, M. (2018). A hybrid geometric spatial image representation for scene classification. PloS one, 13(9), e0203339.
4. Czech, M., Le Moan, S., Hernández-Andrés, J., & Müller, B. (2024). Estimation of daylight spectral power distribution from uncalibrated hyperspectral radiance images. Optics Express, 32(6), 10392-10407.
5. Pipitone, R. N., & DiMattina, C. (2020). Object clusters or spectral energy? Assessing the relative contributions of image phase and amplitude spectra to trypophobia. Frontiers in psychology, 11, 1847.
6. Lin, S., Sprague, T., & Singh, A. K. (2022). Mind reader: Reconstructing complex images from brain activities. Advances in Neural Information Processing Systems, 35, 29624-29636.
7. Snow, J. C., & Culham, J. C. (2021). The treachery of images: how realism influences brain and behavior. Trends in Cognitive Sciences, 25(6), 506-519.
8. Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., ... & Gao, M. (2023). Techniques and challenges of image segmentation: A review. Electronics, 12(5), 1199.
9. Su, T., & Zhang, S. (2017). Local and global evaluation for remote sensing image segmentation. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 256-276.
10. Cai, Z., Fan, Y., Zhu, M., & Fang, T. (2024). Ultra-Lightweight Network for Medical Image Segmentation Inspired by Bio-Visual Interaction. IEEE Transactions on Circuits and Systems for Video Technology.
11. Zhang, X., Zhu, Y., Chen, L., Duan, P., & Zhou, M. (2024). Augmented reality navigation method based on image segmentation and sensor tracking registration technology. Scientific Reports, 14(1), 15281.
12. Zhan, X., Liu, J., Long, H., Zhu, J., Tang, H., Gou, F., & Wu, J. (2023). An intelligent auxiliary framework for bone malignant tumor lesion segmentation in medical image analysis. Diagnostics, 13(2), 223.
13. Wu, J., Yang, S., Gou, F., Zhou, Z., Xie, P., Xu, N., & Dai, Z. (2022). Intelligent segmentation medical assistance system for MRI images of osteosarcoma in developing countries. Computational and mathematical methods in medicine, 2022(1), 7703583.
14. Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R., & Mijic, A. (2020). Image segmentation methods for flood monitoring system. Water, 12(6), 1825.
15. Song, W., Dong, L., Zhao, X., Xia, J., Liu, T., & Shi, Y. (2022). Review of Nodule Mineral Image Segmentation Algorithms for Deep-Sea Mineral Resource Assessment. Computers, Materials & Continua, 73(1).
16. Cai, Z., Hu, Q., Zhang, X., Yang, J., Wei, H., He, Z., ... & Xu, B. (2022). An adaptive image segmentation method with automatic selection of optimal scale for extracting cropland parcels in smallholder farming systems. Remote Sensing, 14(13), 3067.
17. Lemenkova, P. (2020). Object based image segmentation algorithm of SAGA GIS for detecting urban spaces in yaoundé, Cameroon. Central European Journal of Geography and Sustainable Development, 2(2), 38-51.
18. Rasib, M., Butt, M. A., Riaz, F., Sulaiman, A., & Akram, M. (2021). Pixel level segmentation based drivable road region detection and steering angle estimation method for autonomous driving on unstructured roads. IEEE Access, 9, 167855-167867.
19. Agrawal, P., Ratnoo, A., & Ghose, D. (2017). Image Segmentation-Based Unmanned Aerial Vehicle Safe Navigation. Journal of Aerospace Information Systems, 14(7), 391-410.
20. Liu, F., Zhu, J., Lv, B., Yang, L., Sun, W., Dai, Z., ... & Wu, J. (2022). Auxiliary segmentation method of osteosarcoma MRI image based on transformer and U-Net. Computational Intelligence and Neuroscience, 2022(1), 9990092.
21. Ren, T., Wang, H., Feng, H., Xu, C., Liu, G., & Ding, P. (2019). Study on the improved fuzzy clustering algorithm and its application in brain image segmentation. Applied Soft Computing, 81, 105503.
22. Abdel-Basset, M., Chang, V., & Mohamed, R. (2021). A novel equilibrium optimization algorithm for multi-thresholding image segmentation problems. Neural Computing and Applications, 33, 10685-10718.
23. Tong, J., Zhao, Y., Zhang, P., Chen, L., & Jiang, L. (2019). MRI brain tumor segmentation based on texture features and kernel sparse coding. Biomedical Signal Processing and Control, 47, 387-392.
24. Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentation based on U-net: A review. Journal of Imaging Science & Technology, 64(2).
25. Khouy, M., Jabrane, Y., Ameur, M., & Hajjam El Hassani, A. (2023). Medical image segmentation using automatic optimized U-Net architecture based on genetic algorithm. Journal of Personalized Medicine, 13(9), 1298.
26. Vengalil, S. K., Krishnamurthy, B., & Sinha, N. (2023). Simultaneous segmentation of multiple structures in fundal images using multi-tasking deep neural networks. Frontiers in Signal Processing, 2, 936875.

27. Rebouças Filho, P. P., da Silva Barros, A. C., Almeida, J. S., Rodrigues, J. P. C., & de Albuquerque, V. H. C. (2019). A new effective and powerful medical image segmentation algorithm based on optimum path snakes. Applied Soft Computing, 76, 649-670.
28. Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., & Xing, X. (2019). Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. PloS one, 14(4), e0215676.
29. Li, H. A., Fan, J., Hua, Q., Li, X., Wen, Z., & Yang, M. (2022). Biomedical sensor image segmentation algorithm based on improved fully convolutional network. Measurement, 197, 111307.
30. Li, Y., Si, Y., Tong, Z., He, L., Zhang, J., Luo, S., & Gong, Y. (2022). MQANet: Multi-Task Quadruple Attention Network of Multi-Object Semantic Segmentation from Remote Sensing Images. Remote Sensing, 14(24), 6256.
31. Sheng, J., Lv, G., Wang, Z., & Feng, Q. (2022). SRNet: Sparse representation-based network for image denoising. Digital Signal Processing, 130, 103702.
32. Peng, J., Sun, W., Li, H. C., Li, W., Meng, X., Ge, C., & Du, Q. (2021). Low-rank and sparse representation for hyperspectral image processing: A review. IEEE Geoscience and Remote Sensing Magazine, 10(1), 10-43.
33. Yang, M., Zhao, W., Xu, W., Feng, Y., Zhao, Z., Chen, X., & Lei, K. (2018). Multitask learning for cross-domain image captioning. IEEE Transactions on Multimedia, 21(4), 1047-1061.
34. Liu, W., Zhou, C., Li, Z., & Hu, Z. (2020). Patch-driven tongue image segmentation using sparse representation. IEEE Access, 8, 41372-41383.
35. Park, S., Jeong, W., & Moon, Y. S. (2020). X-ray image segmentation using multi-task learning. KSII Transactions on Internet and Information Systems (TIIS), 14(3), 1104-1120.
36. PetcharapornYodjai, PoomKumam, JuanMartínezMoreno & WachirapongJirakitpuwapat. (2024). Image inpainting via modified exemplar-based inpainting with two-stage structure tensor and image sparse representation. Mathematical Methods in the Applied Sciences,47(11),9027-9045.
37. Guo Jianzhong,Cao Cong,Shi Dehui,Chen Jing,Zhang Shuai,Huo Xiaohu... & Guo Min. (2021). Matching Pursuit Algorithm for Decoding of Binary LDPC Codes. Wireless Communications and Mobile Computing,2021.
38. Raphael Hartner,Martin Kozek & Stefan Jakubek. (2025). Multi-task learning with state propagation for quality forecasts in polymer extrusion lines. Journal of Intelligent Manufacturing,(prepublish),1-15.