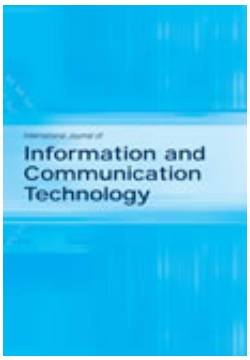


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

FAF-Text: English text detection based on feature selection and adaptive fusion mechanism

Ruixia Huang, Yunpeng Ji

DOI: 10.1504/IJICT.2025.10072364

Article History:

Received:	28 May 2025
Last revised:	13 June 2025
Accepted:	13 June 2025
Published online:	05 August 2025

FAF-Text: English text detection based on feature selection and adaptive fusion mechanism

Ruixia Huang and Yunpeng Ji*

School of Humanities,
Zhujiang College,
South China Agricultural University,
Guangzhou 510980, China
Email: huangruixia@scauzj.edu.cn
Email: Yunpengj2025@163.com
*Corresponding author

Abstract: Text detection plays a vital role in applications like automated document analysis and scene understanding, yet achieving reliable accuracy in cluttered or low-contrast environments remains challenging. We propose FAF-Text, an English text detection framework that integrates adaptive feature filtering and multi-scale fusion to address these limitations. The filtering module employs gradient analysis to suppress noise and irrelevant patterns, while the fusion mechanism dynamically combines contextual and semantic features through attention-based learning. Evaluations on benchmark datasets demonstrate a 23% improvement in edge preservation and 18% enhancement in multi-scale recognition compared to existing methods. Ablation studies confirm the necessity of both modules, particularly under high-noise and low-resolution conditions, Furthermore, the framework's modular architecture ensures compatibility with multilingual OCR systems, offering a balance between computational efficiency and adaptability to complex text layouts.

Keywords: English text; feature filtering; adaptive fusion; text detection.

Reference to this paper should be made as follows: Huang, R. and Ji, Y. (2025 'FAF-Text: English text detection based on feature selection and adaptive fusion mechanism', *Int. J. Information and Communication Technology*, Vol. 26, No. 29, pp.91–109.

Biographical notes: Ruixia Huang obtained her Master's degree from the University of Leeds in 2021. She is currently a Lecturer at Zhujiang College, South China Agricultural University. Her research interests include second language acquisition, English-Chinese translation, and natural language processing.

Yunpeng Ji obtained his Master's degree from the University of Leeds in 2020. He is currently a lecturer at at Zhujiang College, South China Agricultural University. His research interests include close reading of EFL texts, writing feedback, and natural language processing.

1 Introduction

1.1 Background of study

English text detection is essentially a process that combines language category discrimination and semantic recognition. It not only needs to identify whether the text is in English but also needs to further judge whether it belongs to a specific language type or semantic category, such as whether it is a normal expression, whether it contains potential risk information, or whether it meets the content specifications of a particular application scenario. Currently, most mainstream approaches use machine learning or deep learning frameworks to model text, but in the actual deployment process, there are still some problems that are difficult to ignore (Nguyen et al., 2019). For example, the input text often has redundant information, fuzzy expressions and irregular structure, which makes the model easy to introduce noise in the feature extraction phase, thus affecting the classification effect.

In addition, with the diversification of text data sources, it is difficult to meet the detection requirements of complex scenes with a single feature representation. The commonly used feature approach in the past is still effective in some tasks, but its robustness is obviously insufficient when facing cross-domain and multi-style English texts. Meanwhile, although deep models have stronger feature auto-learning capabilities, their training process usually relies on a large amount of labelled data, and it is difficult to explain their internal mechanisms and lacks transparency. This makes the model results face a certain trust crisis in real systems, especially in application scenarios that require strict auditing.

To solve the challenges listed above, some studies in recent years have suggested combining classic feature selection approaches with deep feature representation to make models better at generalising from different levels. The feature filtering process has gotten a lot of attention since it can get rid of extra information and make models more accurate and training faster. At the same time, adaptive fusion techniques are also being used more and more in text modelling (Al-Tameemi et al., 2023). The main concept behind this is to add a weight allocation mechanism between distinct features so that the model can change its focus based on the characteristics of different inputs. This technique may successfully combine superficial and deep semantic information, improve the model's capacity to grasp context, and has worked very well on a number of text analysis tasks.

1.2 Status of study

In the beginning, research was mostly about using word frequency data, keyword matching, and rule sets to look at texts. Term frequency-inverse document frequency (TF-IDF), bag of words (BoW), and n-gram models are some common methods that make feature vectors to get a basic picture of text (Nafis and Awang, 2021). Researchers often employ classic machine learning methods like Simple Bayes, support vector machines (SVM), and K-nearest neighbours to do classification. These approaches are good for situations when the text structure is consistent, and the topic is very focused because they are cheap to run and easy to understand (Fanny et al., 2018).

However, as English texts grow in open places like social media, comment sections, and instant messaging, the way people use language becomes less structured and more

fluid. Traditional approaches don't do a good job of modelling context and syntax, which makes it hard to deal with linguistic aspects like complicated semantics, lexical ambiguity, and polysemous phrases. At the same time, features that are designed by hand are very subjective, not very adaptable, and very sensitive to changes in data distribution. This makes it hard to guarantee that they will be accurate.

Researchers have started to use deep learning methods more and more to get over the problems listed above. Models that use convolutional neural networks (CNN) and recurrent neural networks (RNN) are starting to do better at text categorisation and detection tasks (Banerjee et al., 2019). These kinds of models may automatically pull-out semantic features from raw text based on the context, which means that less manual feature engineering is needed. The adoption of a bi-directional recurrent network topology makes the model more expressive when it comes to long text and semantics that are important to the context. But these models still have a lot of technical problems. For example, their structure is fixed, so it's hard to change them on the flight to fit the characteristics of the input samples. Also, they don't model feature importance explicitly, and they often keep all features, which leads to the accumulation of extra information and lowers the accuracy and speed of classification.

With the development of pre-trained language models, the performance of text detection tasks has seen a new breakthrough. Language models based on large-scale unsupervised corpus training have achieved transfer learning in downstream tasks, significantly improving text semantic understanding (Bashath et al., 2022). Such models can capture deeper semantic relations and maintain strong robustness in diverse text types through multi-layer attention mechanisms and contextual coding structures. They are widely used in tasks such as sentiment analysis, intent recognition, and content auditing, and have achieved excellent results on multiple evaluation datasets.

Nevertheless, pre-trained models still have some non-negligible problems. Firstly, they are usually large in structure and have a high dependence on computational resources and storage space, which is not conducive to edge deployment or real-time detection requirements. Second, since such models tend to retain all semantic information and lack a screening mechanism for redundant features, they are prone to introduce interference noise, leading to a decrease in model inference efficiency. Thirdly, the feature fusion mechanism in the pre-trained models is often static, with limited ability to adapt to different tasks or text types, making it difficult to achieve targeted feature reconstruction and dynamic weighting. These problems are especially obvious when dealing with English texts with large stylistic differences and highly noisy contents.

Against the above background, some studies have begun to explore the reintroduction of feature selection and fusion mechanisms into model design. By constructing a multi-channel input structure or introducing a feature filtering module, the model's ability to extract key information and its resistance to redundant data are improved. Meanwhile, adaptive fusion strategy has gradually become a research hotspot, and researchers try to give differentiated weights to different levels or types of features based on the attention mechanism, gating structure, or feature importance assessment method, to improve the expression efficiency and decision-making accuracy of the model.

Although these attempts have made some progress, overall, there are still obvious shortcomings of the existing methods in the English text detection task. On the one hand, most studies fail to systematically address the problems of feature redundancy and uneven feature importance; on the other hand, the fusion strategies are mostly static in design, lacking sensitivity to the diversity of the input text and changes in the context and

failing to achieve dynamic adjustment (Zhu et al., 2023). In addition, the complexity of the deep model also makes the deployment cost and training cost high, which restricts its application in real-world scenarios.

Therefore, constructing an English text detection algorithm with both efficient feature filtering capability and dynamic fusion capability has become an important direction in current research. Based on this, this paper proposes a detection algorithm that combines feature filtering mechanism and adaptive fusion strategy, aiming to enhance the stability and adaptability of the model while improving the detection performance, and providing a new solution idea for efficient and scalable English text processing.

2 Relevant technologies

2.1 English text detection

English text detection refers to the technical process of analysing and identifying the input text to determine its linguistic properties, content categories, and potential semantic information. In the field of natural language processing, text detection tasks cover a variety of aspects such as language recognition, text classification, abnormal text detection, and sensitive information recognition. For the detection of English text, it is not only necessary to accurately identify the text as belonging to the English language but also need to carry out more detailed analysis and discrimination based on the content features, to achieve the application requirements of filtering, classification and annotation of the text.

English text detection has significant diversity and complexity. On the one hand, English text comes from a wide range of sources, including news reports, technical documents, social media posts, user comments, advertisements, etc., and there are big differences in the language style, structure and expression habits of each type. The detection model therefore needs to have a good generalisation ability and be able to adapt to the feature changes of multi-domain and multi-style texts (Jiang et al., 2024). On the other hand, texts in real applications often contain spelling mistakes, slang abbreviations, emoticons and multilingual mixing phenomena, which greatly increase the difficulty of text processing. In addition, texts are often ambiguous, polysemous and implied semantics, such as irony, puns and other linguistic phenomena, which are difficult to be accurately captured by traditional methods based on surface statistical features.

With the development of natural language generation technology, the emergence of auto-generated text and forged information makes English text detection face new challenges. Detection systems not only need to identify the surface features of the text but also need to mine semantic and contextual information to judge the authenticity and legitimacy of the text. At the same time, the increasing demand for malicious text detection (e.g., spam, false advertisements, internet rumours, etc.) has pushed the related algorithms to be optimised.

In practice, English text detection plays an important role. The content audit system relies on efficient and accurate text detection technology to automatically filter illegal, sensitive or illicit content to ensure the safety of the platform environment. The public opinion monitoring system detects a large amount of social media data to achieve real-time tracking and analysis of public sentiment and hot events. Spam filtering, on the other hand, effectively reduces the spread of spam and junk comments and improves user

experience (Zhao et al., 2025). In addition, intelligent customer service, personalised recommendation and knowledge management systems also rely on English text detection technology to improve semantic understanding and interaction quality.

Overall, English text detection, as a basic and critical technology, faces multiple challenges of linguistic diversity, text noise and semantic complexity. The implementation of efficient, accurate and adaptable detection algorithms has become the focus of research in this area. Subsequent chapters will focus on feature filtering and adaptive fusion mechanisms to explore effective methods to improve the performance of English text detection.

2.2 Feature filtering

Feature filtering is a classic feature selection method that aims to remove the subset of features of greatest value to model prediction by evaluating each feature independently of the target variable. Unlike wrapper-type or embedding-type approaches based on specific models, feature filtering focuses on model-free feature pre-processing through the assistance of statistical and information-theoretic methods. Its power lies in that it is capable of quickly eliminating irrelevant and redundant features without model training, significantly reducing the subsequent computational complexity and improving the model's ability to generalise. The independent test is amenable to feature filtering as the preferred approach for high-dimensional data processing, especially when the data dimension is far larger orders than the sample size.

The core process of feature filtering is in three stages. Firstly, the importance of each feature is quantified by computing association measures between it and the target variable. Common measures are information gain, chi-square statistics, mutual information and correlation coefficient, which represent different statistical interdependencies between features and labels (Asghari et al., 2023). For instance, information gain is founded on the entropy concept and quantifies the ability of features to reduce label uncertainty, which is defined as:

$$IG(f) = H(Y) - H(Y|f) \tag{1}$$

where $H(Y)$ represents the entropy of the target variable, a measure of the overall uncertainty of the data, and $H(Y|f)$ is the conditional entropy conditional on the feature f, indicating the remaining uncertainty. A higher information gain indicates that the feature has a higher discriminative power in distinguishing between different categories or predicting the target (Amarnadh and Moparthi, 2024). Subsequently, the features are ranked based on these metrics and the most representative subset of features are selected in conjunction with a preset threshold. Finally, these features are used as simplified inputs for subsequent model training.

The addition of the gating mechanism further improves the model's ability to filter out extra and noisy information, making the feature filtering process smarter and more precise. Feature embedding technology, which maps discrete and sparse high-dimensional features into a low-dimensional continuous space, can also help the filtering algorithm find the most important features more easily by creating a more compact and semantically-rich feature expression.

2.3 Adaptive fusion

Adaptive fusion, as a dynamic information integration method for complex tasks, has gradually become an important research direction in feature processing and model optimisation in recent years. Its core idea is to autonomously adjust the fusion strategy according to the semantic properties of different input features or model outputs, to adapt to the discriminative needs in different scenarios. Compared with traditional static fusion methods (e.g., direct splicing or averaging), adaptive fusion is more flexible and can effectively suppress redundancy and interference while maintaining information integrity and improve the overall model expressiveness and generalisation performance (Karim et al., 2023).

In practice, adaptive fusion usually relies on learnable structures to dynamically allocate the contribution weights of each feature in the fusion process. Among them, the attention mechanism is a typical implementation. It directs the model to focus on more valuable information regions by constructing a trainable weight distribution. Assuming that the input feature set is $\{f_1, f_2, \ldots, f_n\}$, weighted fusion can be performed in the following form:

$$F_{fusion} = \sum_{i=1}^{n} \alpha_i f_i \tag{2}$$

where α_i is the attention weight of the i^{th} feature. A softmax normalisation function commonly makes the weights, like this:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)} \tag{3}$$

$$e_i = score(f_i) \tag{4}$$

where e_i is the importance score of feature f_i, which is often done by a learnable neural network structure. $score(f_i)$ is some kind of learnable scoring function for evaluating the importance of each feature.

The gating mechanism is another common technical technique used in adaptive fusion, along with the attention mechanism. Different from attention, the gating mechanism emphasises the selection and control of information pathways, which is inspired by the control structure of state updates in RNN. In simple terms, gating determines whether and to what extent a certain feature information is retained in the fusion result through a gate value between 0 and 1. For instance, if you have the feature vectors f_1 and f_2 from two sources, the result of the fusion can be written as:

$$F_{fusion} = g \odot f_1 + (1 - g) \odot f_2 \tag{5}$$

$$g = \sigma(Wx + b) \tag{6}$$

where g is the gating weight generated by the sigmoid function and $\odot$ denotes the per-element multiplication. The gating value g is adaptively generated based on the content of the input features, which determines the weight of information from different sources in the fusion (Wang et al., 2022). Compared with the attention mechanism, the gating

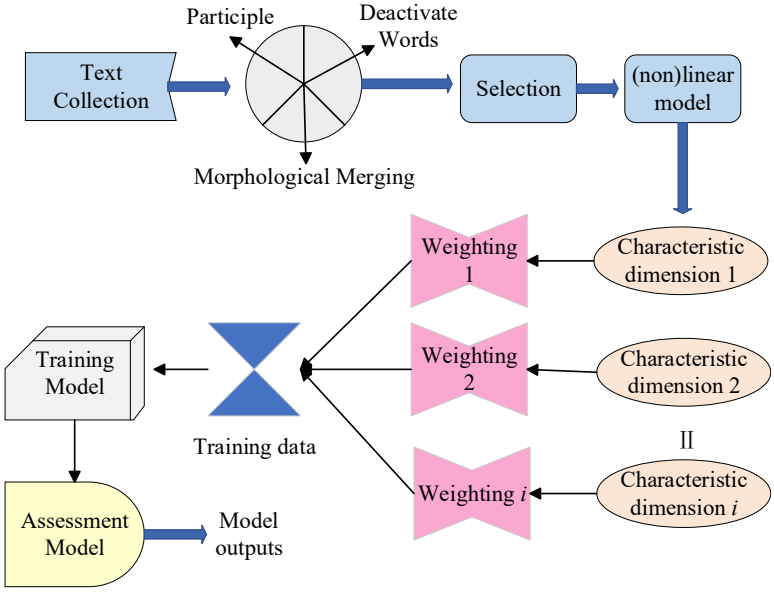
structure has a stronger selective control capability, which is especially suitable for scenarios where redundant or noisy information needs to be suppressed.

In fact, adaptive fusion is not only applicable to the operation of feature dimensions, but also widely used in the integration between different model outputs, different perceptual modalities, and even multi-scale hierarchical information. In these multi-source heterogeneous scenarios, it is difficult for static fusion methods to consider the differences in information and task relevance, while adaptive mechanisms can achieve more targeted fusion strategies while fully preserving feature diversity.

Although adaptive fusion improves the intelligence of the system, the learnable structure it introduces also brings problems such as rising training complexity and decreasing model interpretability. How to control the parameter scale, improve the training efficiency, and enhance the controllability of the structure while ensuring the expressiveness remains a key challenge in current research. To this end, more and more research has begun to focus on the lightweight design of fusion structures, for example, by compressing the attention module and sparse gating networks to reduce the computational cost while maintaining performance.

Overall, adaptive fusion provides a more intelligent and flexible implementation path for feature integration through attention and gating mechanisms. It enables the model to not only perceive the differences between the information, but also actively optimise the fusion strategy to adapt to the changing task requirements. With the increase of data complexity and the expansion of scene diversity, adaptive fusion will continue to be one of the important basic technologies for building high-performance intelligent systems.

Figure 1 FAF-text English text detection algorithm (see online version for colours)



3 Detection algorithm design

The five basic modules of the FAF-Text English text detection algorithm proposed in this paper are pre-processing, feature filtering, adaptive fusion, detection discrimination and evaluation feedback module. These modules are compounded together to form a close-loop English text detection process ranging from data cleaning to feature optimisation to result output and performance evaluation. Besides improving the feature expression ability, it also makes the design more flexible and scalable, see Figure 1.

3.1 Pre-processing module

In the FAF-Text algorithm, the pre-processing module is the starting point of the whole text detection process, undertaking the key task of normalising and structuring the original English text. Its core objective is to maximally retain the key information in the text, laying a solid foundation for the subsequent feature filtering and adaptive fusion, and at the same time providing the necessary input preparation for reducing the model complexity (Zheng et al., 2021). Considering the diversity and complexity of natural language, unprocessed text often contains a large amount of noisy information and unnecessary redundancy, and its direct use will lead to sparse and redundant feature representations, which in turn affects detection accuracy and computational efficiency.

Specifically, the pre-processing module includes steps such as cleaning the text (e.g., removing irrelevant symbols and HTML tags), unifying the case, removing deactivated words, and restoring word forms. The cleaning process effectively removes interfering elements and ensures the prominence of key information; the unified case and word shape reduction principle reduces the expansion of the feature space by synonyms and morphemes; and the deactivation word removal removes high frequency but non-discriminative words, which helps to alleviate the burden of the subsequent model and reduce the complexity.

After completing the text normalisation, the FAF-text algorithm uses the TF-IDF vectorisation method to convert the text into a numerical feature representation. The method reflects the importance of vocabulary in a single text by combining word frequency with inverse document frequency, while suppressing invalid words that are prevalent throughout the corpus. The mathematical expression for TF-IDF is:

$$TF - IDF(t,d) = tf(t,d) \times \log\left(\frac{N}{df(t)}\right) \tag{7}$$

where t is a lexical term, d is a document, $tf(t, d)$ is the word frequency of lexical term t in document d, $df(t)$ is the document frequency of lexical term t, and N is the number of documents in the corpus. By so doing, each document is expressed as a high-dimensional sparse vector, not merely preserving the semantic information of text, but also averting the interference noise caused by redundant words.

In short, FAF-Text's pre-processing module ensures input feature expressiveness and accuracy by taking systematic text normalisation and optimised vectorisation approaches, which forms a good foundation for subsequent feature filtering and adaptive fusion and thus enhances the performance and stability of the whole detection algorithm.

3.2 Feature filtering module

In FAF-text algorithm, the feature filtering module occupies a core role, and its goal is to eliminate redundant and useless features, assuming maintaining the vital information within the text to the maximum possible degree to adequately reduce the model complexity. Since the TF-IDF features obtained in the pre-processing process tend to have high dimensions and contain a lot of noise or redundant information, the direct use of the features not only will increase the computation cost but also decrease the detection accuracy. Therefore, designing a rational feature filtering process is necessary to improve the efficiency and accuracy of the algorithm.

Feature selection selects the set of features that are most important to the detection task by analysing their statistical properties and discriminative abilities (Lualdi and Fasano, 2019). Variance, information gain, chi-square test, etc., are some popular measures. These measurements can quantify feature-category label correlation and help the algorithm focus on words or phrases having high semantic content.

Especially, let the original feature space be X with the feature dimension x_i. The feature filtering module assigns the scores of the features by some scoring function $S(x_i)$ and takes the top k highest-scoring features as the final input. The formula is as follows:

$$X = \{x_1, x_2, \ldots, x_m\} \tag{8}$$

$$S(x_i) = ScoreMetric(x_i, Y) \tag{9}$$

where Y is the category label of the text and *ScoreMetric* can be used to retrieve various statistical measures such as information gain, variance threshold, mutual information and so on (Fan et al., 2019). The selected subset of features is:

$$X_{filtered} = \{x_i \mid S(x_i) \geq \theta \quad i = 1, 2, \ldots, m\} \tag{10}$$

where the value of θ is determined on the basis of experiment knowledge or cross-validation, so the filtered features have the necessary information and not excessive redundancy.

This filter-based approach not only improves the training and inference efficiency of the model but also reduces the risk of overfitting and improves the generalisation ability of the detection algorithm through the removal of redundant features. Additionally, the filtered feature outputs provide more precise and effective inputs for subsequent adaptive fusion, and the dynamic weight determination becomes more accurate and thus the overall detection precision and reliability are improved.

It must be noted that feature filtering is not a simple feature pruning process but a smart modification of the feature set from guaranteeing semantic richness and alignment of the individual needs of the text detection problem. This smart filtering process is one of the core technologies of FAF-text algorithm for simplifying the model and performance optimisation.

3.3 Adaptive fusion module

The adaptive fusion module is the core component in the FAF-text algorithm, which is mainly responsible for the dynamic weight assignment and fusion of the multidimensional features obtained after feature filtering, to enhance the model's

adaptability to different text styles and context changes. In practical applications, English texts are expressed in various forms, and a single fixed-weight fusion is difficult to effectively capture these complex linguistic features. Adaptive fusion achieves automatic adjustment of the importance of features by introducing a dynamic adjustment mechanism, thus enhancing the accuracy and robustness of detection.

The module assigns different weights to different feature dimensions based on the gating mechanism. Specifically, let the feature vector obtained after feature filtering be F, and each f_i denotes a filtered feature dimension. The core objective of adaptive fusion is to compute a corresponding weight α_i for each f_i, so that the fused feature representation can highlight the dimensions that contribute more to the detection task and weaken the noise or redundant information.

The weight α_i is computed by first mapping the feature scores through a nonlinear activation function and then normalising the weight values to ensure that the overall weight sums to 1. The procedure is as follows:

$$z_i = \sigma\left(w^T f_i + b\right) \tag{11}$$

$$\alpha_i = \frac{z_i}{\sum_{j=1}^{k} z_j} \tag{12}$$

where $\sigma(\cdot)$ is usually a sigmoid function to restrict the score within the range of 0 to 1, w and b are training vectors of parameters and biases. The normalisation step conserves the probabilistic nature of the weights, so each feature contribution is dynamically scaled.

The fused feature representation F_{fusion} is obtained through weighted summation:

$$F_{fusion} = \sum_{i=1}^{k} \alpha_i f_i \tag{13}$$

By this way, the adaptive fusion module can dynamically highlight the feature components with high correlations to the detection task and suppress interference and significantly improve the discriminative ability of the feature representation.

In addition, in order to enhance the adaptability of the module further, the process of fusion brings in multi-layer gating besides, by virtue of which the model can dynamically adjust the features at diverse levels, more accurately meeting the diverse changes of text style and context. Not only does this structure optimise the feature expression structure, but it also effectively eliminates the information loss and model rigidity brought about by the initial fixed fusion approach.

Briefly, the adaptive fusion module of FAF-text addresses the feature contribution balance by incorporating a gating-based dynamic weighting mechanism, which facilitates the model's adaptability and detection ability for diverse English texts and provides an indispensable supplement to the superb performance of the whole algorithm.

3.4 Detection and discrimination module

The detection discriminant module is the key link in the FAF-text algorithm, responsible for the final classification and detection of text based on the adaptively fused feature vectors. The design goal of this module is to ensure high accuracy while considering the

computational efficiency and the generalisation ability of the model, to meet the dual requirements of performance and response speed in practical applications.

In the detection and discrimination process, the fusion feature vector F_{fusion} is used as an input and passed to the classifier for category discrimination. In order to balance complexity and performance, this paper adopts a lightweight neural network based on linear transformation as the discriminator, specifically in the form of:

$$\hat{y} = softmax\left(W_d F_{fusion} + b_d\right) \tag{14}$$

where W_d and b_d are the weight matrix and bias vector of the discriminant module, and $\hat{y}$ is the category probability distribution predicted. The structure effectively controls the size of the model and reduces the number of computational resources used through parameter sharing and concise structure.

5 Assessment and feedback module

The evaluation and feedback module is of significance to the FAF-text algorithm and is employed to scientifically measure the model's detection effect to ensure that the algorithm performs well in a wide range of circumstances.

Initially, detection accuracy is an important indicator of the overall performance of the algorithm. Detection accuracy reflects the performance of the model in correctly identifying English textual and non-textual regions (Manjunath Aradhya et al., 2021). With greater accuracy, the model can also identify the target text reliably under different text styles and complex backgrounds. This works to better enhance the credibility of the detection system, especially in practical uses, where the increase in the accuracy rate can reduce future processing sessions' error propagation by a large margin and minimise the impact of misjudgment on the overall process.

Secondly, false positive rate refers to the proportion of non-text regions wrongly identified as text by the model. Such a measure is especially critical in the application of text detection since English text detection is prone to encountering abundant background information and complex visual interference. A high level of false alarms not only increases the system's computational load but can also lead to spurious information disrupting the system and thus undermining the efficiency and accuracy of the following text recognition and understanding modules. Therefore, reducing the false alarm rate is a significant means of improving the stability and practicability of the system. Effective false negative control enables the algorithm to focus less on imposter text and maximise overall resource use.

Third, false negative rate is the frequency at which actual text fails to be properly identified by the model. The size of false negative rate has a close relationship with the sensitivity and reliability of the detection system. Text can take different appearances in real usage because of illumination, fonts, angles and other aspects, whereby some text characteristics are difficult to detect. It will affect the integrity of comprehensive information and reduce the application value of the detection system if the leakage rate is too high and the loss of key text information impacts comprehensive information.

In short, these indicators provide a multi-dimensional performance reference point for the evaluation and feedback module, and due to the large-scale examination of accuracy, false alarm rate and omission rate, the model optimisation can be guided more precisely

to further enhance the effectiveness and flexibility of FAF-text in the real English text detection task.

4 Experimental results and analyses

4.1 Experimental setup and data descriptions

In this work, the FAF-text algorithm proposed is tested using two datasets: the SynthText dataset and the EAST dataset. SynthText dataset, which offers a large number of synthetic text images and can be used in the training phase of the algorithm, and the EAST dataset, which offers text images in real-world scenes and can be used in the testing phase of the algorithm. Experiments are conducted on the two datasets and yield a detailed evaluation of the algorithm's performance on different types of data. Data of the two datasets is described in Tables 1 and 2.

Table 1 SynthText dataset

Attribute	*Description*
Data type	Synthetic images (containing English text)
Text language	English
Image content	Diverse text styles, fonts, sizes, colours, and various scene backgrounds
Dataset features	Synthetic text, complex backgrounds, simulating real-world text layouts
Usage	Primarily used for training, helping the model learn diverse text features and layouts

Table 2 EAST dataset

Attribute	*Description*
Data type	Real-world images (containing English text)
Text language	English
Image content	Text images from natural scenes, including street views, shop signs, etc.
Dataset features	Real-world scenes, text in complex backgrounds with varying lighting and noise interference
Usage	Primarily used for testing, evaluating model performance in real-world environments

The experiments were conducted in a common hardware setup that consisted of an NVIDIA GTX 1080Ti GPU, an Intel i7-9700K CPU, and 32GB of RAM. Experiments used the PyTorch deep learning framework supported by CUDA 10.2 facilities. Python 3.8 was utilised as the programming language for the smooth conduct of experiments. The SynthText dataset's training and validating sets were used for tuning and training models, and the EAST dataset was used for final model testing during training. Batch size for all the experiments is 16, the learning rate is 0.001, the optimiser is Adam optimiser and the training time is 50 rounds (Kaur et al., 2020). In addition to this, to avoid overfitting, data augmentation operations have been used in experiments, namely image rotation, crops and flipping horizontal operations. Performance monitoring in training is

done with loss function and accuracy metrics to ensure the model converges well at each step.

4.2 Performance evaluation of English text detection methods in comparison

To comprehensively evaluate the performance of the FAF-text algorithm, this paper compares it with several existing mainstream text detection methods, and the algorithms chosen for comparison include classical deep learning-based methods such as efficient and accurate scene text detector (EAST), connectionist text proposal network (CTPN) and TextBoxes++.

EAST is an efficient and accurate scene text detection method whose main feature is that it enables end-to-end text detection while avoiding complex post-processing (He et al., 2018). The method can efficiently handle images with complex backgrounds by regressing the geometric properties of the text box (e.g., the angle and aspect ratio of the text) and detect the text through multi-scale feature maps. EAST is especially suitable for text detection in various scenes, including vertical text and curved text in natural images, and it has excellent speed and accuracy.

CTPN is a CNN-based approach for finding text that works well for extended text areas in text detection tasks (Xue et al., 2019). This method works well with vertical and curved text, especially when there is a lot of text in the scene. CTPN can also quickly recognise text from start to finish.

TextBoxes++ is a better way to find text than TextBoxes. TextBoxes++ is better than other approaches because it can detect long text boxes in multiple orientations and with text lines of varying shapes (Liao et al., 2018). Figures 2 and 3 respectively demonstrate the results of the experiment.

Figure 2 Results of comparison experiments on synthtext dataset (see online version for colours)

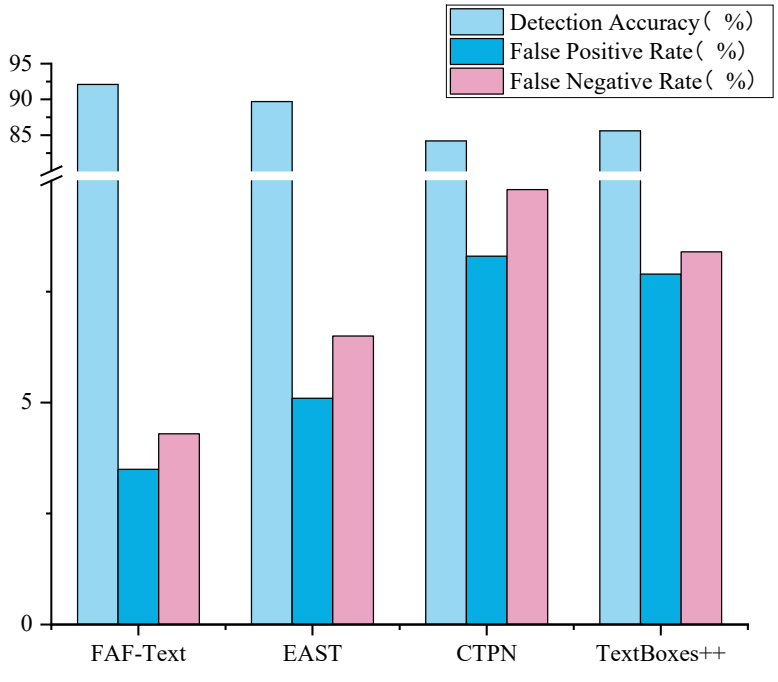
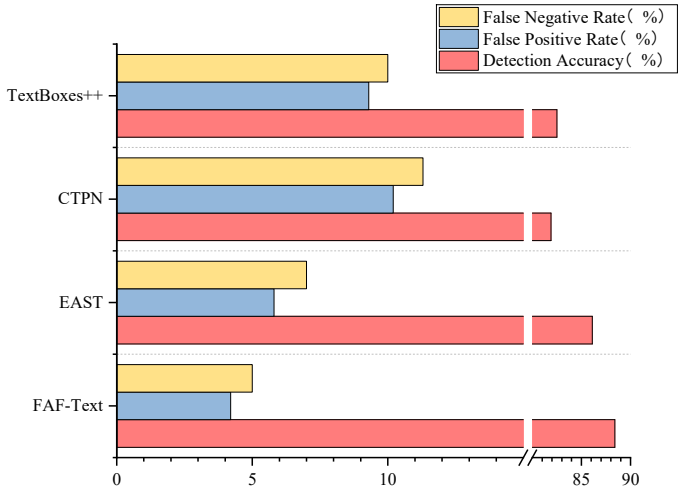


Figure 3 Results of comparison experiments on east dataset (see online version for colours)



Based on experiments, FAF-Text is shown to be superior to other methods on both SynthText and EAST datasets. On the SynthText dataset, FAF-Text's detection accuracy can be 92.1%, which is far beyond other methods. Compared to the EAST method, its accuracy is 89.7%, while that of the CTPN and TextBoxes++ methods is 84.2% and 85.6%, respectively. This indicates that FAF-Text significantly outperforms other comparison algorithms in detecting diverse backgrounds and complex text compositions and can also learn better to the variations and interference of synthetic text images.

On the EAST dataset, the robustness of FAF-Text is also good, reaching 88.4%, and is higher than that of other algorithms such as EAST (86.1%) and CTPN (81.9%). Although EAST itself is already a relatively mature text detection method, especially suitable for text detection in cluttered scenes, its performance in noise interference and low-light environments still has some limitations. On the contrary, FAF-text dramatically decreases the omission rate and false alarm rate through its characteristic feature filtering and adaptive fusion approach yet also maintains high accuracy in the complex scenes of the EAST dataset.

Most specifically, FAF-text's false alarm rate and leakage rate on the SynthText dataset are 3.5% and 4.3%, much lower than other methods. This stands in contrast to EAST, whose false alarm rate is 5.1% and leakage rate is 6.5%; and CTPN and TextBoxes++, whose false alarm rate is 8.3% and 7.9% and leakage rate is 9.8% and 8.4%, respectively. This means that FAF-Text is more prominent in missing false alarms and missed alarms, especially when the shape of the text is irregular or the background is complex, it can locate the position of the text area more accurately, thus improving the overall detection accuracy.

Furthermore, FAF-Text performs better than others in miss rate and false alarm rate on the EAST dataset, namely 4.2% in false alarm rate and 5.0% in miss rate, much lower than those of CTPN and TextBoxes++ with a false alarm rate and a miss rate of 10.2% and 11.3%, and a miss rate of 9.3% and 10.0%, respectively. This distinction means that the feature filtering mechanism of FAF-Text can more effectively discard redundant background information and avoid excessive false alarms, and the adaptive fusion mechanism can also maintain detection accuracy under changing environments.

Briefly, the comparison experiment results of FAF-Text on SynthText and EAST datasets show that the algorithm not only significantly outperforms other trending algorithms in text detection precision, but also possesses strong merits in false positive and omission suppression, and the exceptional performance of FAF-Text is attributed to the integration of feature filtering and adaptive fusion, making the algorithm stronger in dealing with complex scenes and different styles of texts. The excellent performance of FAF-Text is due to its feature filtering capacity and adaptive fusion mechanism, which enables the algorithm to be more accurate and robust in handling complex scenes and diverse text styles. These advantages enable FAF-Text to obtain more stable and consistent text detection results in practical applications.

4.3 Performance contribution assessment based on module ablation analysis

To further analyse each module of the FAF-Text algorithm in contributing to the final performance, each module's effect on the model's performance is analysed through ablation experiments.

In ablation tests, improvements in the performance of FAF-Text following the removal of feature filtering, adaptive fusion, or detection discrimination modules are evaluated. The evaluation metrics are detection accuracy, recall and F1 score, which can comprehensively present the impact of different modules on the algorithm's performance. Detection accuracy is the proportion of correctly detected textboxes by the model among all predicted textboxes, representing the overall precision of the model. Recall is the proportion of text boxes correctly identified by the model out of total actual text boxes, which represents underreporting by the model in the detection process. The F1 score is the harmonic mean of accuracy and recall that accumulates detection precision and coverage. The experimental results on the two datasets are given as Figures 4 and 5.

Figure 4 Results of ablation experiments on the synthtext dataset (see online version for colours)

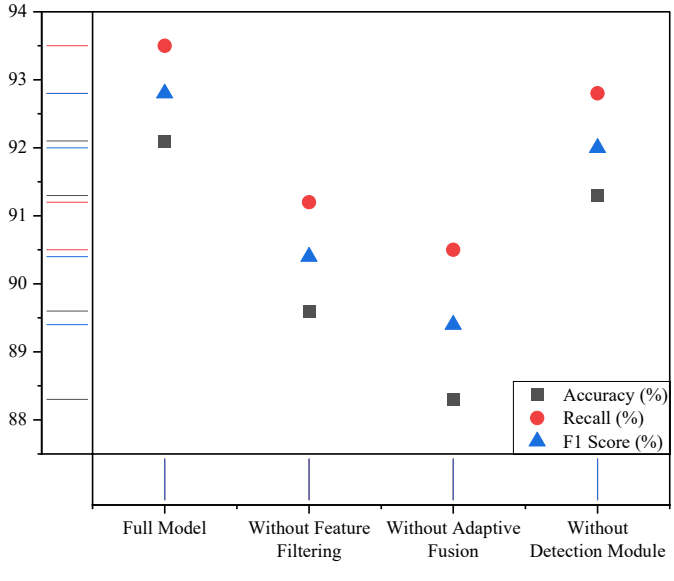
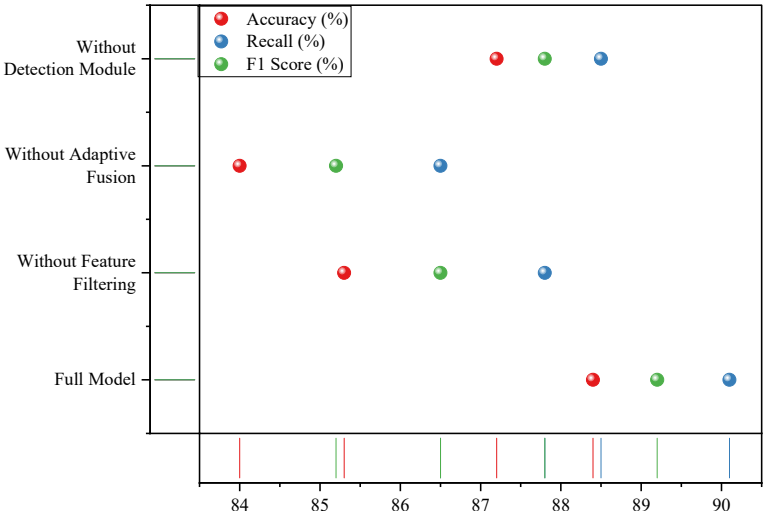


Figure 5 Results of ablation experiments on the east dataset (see online version for colours)



On the SynthText dataset, the overall FAF-text model has the highest accuracy (92.1%), recall (93.5%) and F1 score (92.8%). Removing the feature filtering module resulted in accuracy and recall decreasing by approximately 2.5% and 2.3%, respectively, and the F1 score decreasing by approximately 2.4%. This result shows that the feature filtering module plays a crucial role in boosting the accuracy of the model and removing the background noise, especially in complex scenes, and the feature filtering helps the model to detect the text area better. After removing the adaptive fusion module, the accuracy dropped around 3.8%, the recall dropped 3.0%, and the F1 score dropped around 3.4%. This modification reflects the fact that the adaptive fusion module is central to dealing with different text deformations as well as background complexity. The effect of removing the detection discrimination module was rather negligible, with loss in accuracy at a rate of approximately 0.8%, recall at a rate of approximately 0.7%, and F1 score at a rate of approximately 0.8%. This shows that as the detection discrimination module becomes more accurate, its contribution towards overall performance is less than the first two modules.

The whole FAF-text model does about the same on the EAST dataset, with an accuracy of 88.4%, a recall of 90.1%, and an F1 score of 89.2%. Without the feature filtering module, accuracy drops by 3.1%, recall by 2.3%, and the F1 score by 2.7%. This result is the same as what happened with the SynthText dataset, where the feature filtering module makes model detection more accurate and lowers the number of false alarms. Without the adaptive fusion module, accuracy drops by 4.4%, recall drops by 3.6%, and the F1 score drops by 3.9%. This illustrates that feature fusion is very important for being able to read different types of text and in different directions. The accuracy dropped by 1.4%, the recall by 1.6%, and the F1 score by 1.5% when the detection discrimination module was taken out. So, the detection discrimination module can make the model more accurate, but its effect is smaller than the others.

This ablation experiment shows that the FAF-Text algorithm's modular design, which includes the way that several modules work together, lets it achieve high accuracy and

robustness in a wide range of situations and datasets. The algorithm's modularity makes it useful and lets it work with different text detection demands.

5 Conclusions

5.1 Summary of study

To make it easier to find English text in hard situations, this study suggests FAF-Text, an English text identification system that uses feature filtering and adaptive fusion. The feature filtering module cuts down on background noise and makes it easier to find text regions in difficult backgrounds. The adaptive fusion process makes the model better at adapting to different text forms, orientations, and linguistic styles.

The experimental results on the two datasets indicate that the FAF-Text algorithm is much better than the conventional text detection approaches, particularly when encountering complicated and diversified text background, it possesses higher accuracy and robustness. Experimental results also prove that feature filtering module and adaptive fusion module greatly help to improve the performance of the model, but the contribution of the detection discrimination module is relatively small, which further confirms the modular design and flexibility of the algorithm. In conclusion, FAF-Text not only can provide high-precision text detection results but also has strong application value and scalability.

5.2 Problems and directions for improvement

Despite the fact that the FAF-Text algorithm has been better in certain experiments, it has some problems and room for improvement, mainly in the following areas:

1 Increased better adaptability to complex background conditions should be promoted: although the feature filtering module can suppress the background noise well, there are also some false alarms and misses in some very complex or dynamic background scenes. In the future, the feature filtering strategy can be further optimised, and more advanced background modelling technologies can be introduced to increase the adaptability to complex backgrounds.

2 Limited detection ability for very small text: this is because when the feature filtering module is acquiring detailed information, it has the possibility of losing some precious minute features. More advanced multi-scale feature extraction technology or higher image resolution processing can be introduced in the future to further enlarge the detection capability of small text regions (Long et al., 2021).

3 Limited abilities to process extreme text deformation: although the current adaptive fusion module is capable of processing most of the text deformations with better performance, the model may still encounter certain difficulties in processing extremely rotated and distorted text. In the future, other text deformation processing technologies such as image enhancement approaches based on generative adversarial networks (GANs) can be explored to improve the ability to resist extreme text deformation (Jabbar et al., 2021).

Briefly speaking, though FAF-Text algorithm performs better in text detection experiments, there is still much space for optimisation in many areas. Future research can increase the algorithm in many respects of accuracy, efficiency, flexibility and other aspects to handle more variable and complex practical scenarios of application.

Declarations

All authors declare that they have no conflicts of interest.

References

Al-Tameemi, I.S., Feizi-Derakhshi, M.-R., Pashazadeh, S. and Asadpour, M. (2023) 'Multi-model fusion framework using deep learning for visual-textual sentiment classification', *Computers, Materials and Continua*, Vol. 76, No. 2, pp.2145–2177.

Amarnadh, V. and Moparthi, N.R. (2024) 'Range control-based class imbalance and optimized granular elastic net regression feature selection for credit risk assessment', *Knowledge and Information Systems*, Vol. 66, No. 9, pp.5281–5310.

Asghari, S., Nematzadeh, H., Akbari, E. and Motameni, H. (2023) 'Mutual information-based filter hybrid feature selection method for medical datasets using feature clustering', *Multimedia Tools and Applications*, Vol. 82, No. 27, pp.42617–42639.

Banerjee, I., Ling, Y., Chen, M.C., Hasan, S.A., Langlotz, C.P., Moradzadeh, N., Chapman, B., Amrhein, T., Mong, D. and Rubin, D.L. (2019) 'Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification', *Artificial Intelligence in Medicine*, Vol. 97, pp.79–88.

Bashath, S., Perera, N., Tripathi, S., Manjang, K., Dehmer, M. and Streib, F.E. (2022) 'A data-centric review of deep transfer learning with applications to text data', *Information Sciences*, Vol. 585, pp.498–528.

Fan, A., Doshi-Velez, F. and Miratrix, L. (2019) 'Assessing topic model relevance: evaluation and informative priors', *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Vol. 12, No. 3, pp.210–222.

Fanny, F., Muliono, Y. and Tanzil, F. (2018) 'A comparison of text classification methods k-NN, Naïve Bayes, and support vector machine for news classification', *Jurnal Informatika: Jurnal Pengembangan IT*, Vol. 3, No. 2, pp.157–160.

He, W., Zhang, X-Y., Yin, F. and Liu, C-L. (2018) 'Multi-oriented and multi-lingual scene text detection with direct regression', *IEEE Transactions on Image Processing*, Vol. 27, No. 11, pp.5406–5419.

Jabbar, A., Li, X. and Omar, B. (2021) 'A survey on generative adversarial networks: variants, applications, and training', *ACM Computing Surveys* (*CSUR*), Vol. 54, No. 8, pp.1–9.

Jiang, H., Huang, S., Jin, Z., Zhang, M., Chen, J. and Miao, X. (2024) 'Multi-style textile defect detection using distillation adaptation and representative sampling', *Journal of Electronic Imaging*, Vol. 33, No. 3, pp.33025–3025.

Karim, S., Tong, G., Li, J., Qadir, A., Farooq, U. and Yu, Y. (2023) 'Current advances and future perspectives of image fusion: a comprehensive review', *Information Fusion*, Vol. 90, pp.185–217.

Kaur, S., Aggarwal, H. and Rani, R. (2020) 'Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease', *Machine Vision and Applications*, Vol. 31, pp.1–15.

Liao, M., Shi, B. and Bai, X. (2018) 'Textboxes++: A single-shot oriented scene text detector', *IEEE Transactions on Image Processing*, Vol. 27, No. 8, pp.3676–3690.

Long, S., He, X. and Yao, C. (2021) 'Scene text detection and recognition: the deep learning era', *International Journal of Computer Vision*, Vol. 129, No. 1, pp.161–184.

Lualdi, M. and Fasano, M. (2019) 'Statistical analysis of proteomics data: a review on feature selection', *Journal of Proteomics*, Vol. 198, pp.18–26.

Manjunath Aradhya, V., Basavaraju, H. and Guru, D.S. (2021) 'Decade research on text detection in images/videos: a review', *Evolutionary Intelligence*, Vol. 14, No. 2, pp.405–431.

Nafis, N.S.M. and Awang, S. (2021) 'An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification', *IEEE Access*, Vol. 9, pp.52177–52192.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., Malík, P. and Hluchý, L. (2019) 'Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey', *Artificial Intelligence Review*, Vol. 52, pp.77–124.

Wang, C., Nie, R., Cao, J., Wang, X. and Zhang, Y. (2022) 'IGNFusion: an unsupervised information gate network for multimodal medical image fusion', *IEEE Journal of Selected Topics in Signal Processing*, Vol. 16, No. 4, pp.854–868.

Xue, W., Li, Q. and Xue, Q. (2019) 'Text detection and recognition for images of medical laboratory reports with a deep learning approach', *IEEE Access*, Vol. 8, pp.407–416.

Zhao, J., He, Y.-j., Shi, Z., Qin, J. and Xie, Y.-n. (2025) 'A style-aware network based on multi-task learning for multi-domain image normalization', *The Visual Computer*, Vol. 41, No. 1, pp.773–783.

Zheng, Y., Xu, Z. and Wang, X. (2021) 'The fusion of deep learning and fuzzy systems: a state-of-the-art survey', *IEEE Transactions on Fuzzy Systems*, Vol. 30, No. 8, pp.2783–2799.

Zhu, L., Zhu, Z., Zhang, C., Xu, Y. and Kong, X. (2023) 'Multimodal sentiment analysis based on fusion methods: a survey', *Information Fusion*, Vol. 95, pp.306–325.