World Journal of Information Systems 1(2) 37-52
DOI: https://doi.org/10.17013/wjis.v1i2.12

Article

Identifying Document Patterns with Limited Data: A Multimodal Approach

João Paulo Vieira Costa [1], Bruno Lima Maciel [2] and Li Weingang [3]

[1, 2, 3] Universidade de Brasília - UNB

***** Correspondence: jpcosta1990@gmail.com

(Article history: Received March 20, 2024; Received in revised form April 25, 2024; Accepted June 09, 2024; Available online June 15, 2024)

Abstract: Pattern identification in documents plays a crucial role in the banking sector and others, as it is applied in various operations such as document analysis, information retrieval, and anomaly detection. Challenges arise when traditional training methods require vast labeled datasets, which are costly and time-consuming. This requirement imposes limitations on the generalization and practical applicability of the models developed. This study evaluates the performance of models in a limited data scenario; by leveraging multimodal information and transfer learning, it aims for accurate identification with reduced data requirements. The ability to train efficient models with scarce data proposes a methodological evolution, reducing the need for large volumes of labeled data and promoting more agile and scalable approaches. We compare Few-Shot Learning (FSL) approaches with reference models. The results demonstrate that, even with limited samples, it is possible to achieve performances comparable to established benchmarks. Future research may further explore the behavior of these methods with varied sample sizes and their ability to detect different types of anomalies in unseen document patterns using the Once Learning approach.

Keywords: document pattern; few-shot learning; anomaly detection; once-learning

1. Introduction

In recent years, there has been a significant increase in the use of digital transactions in markets traditionally dominated by physical transactions. According to the Brazilian Federation of Banks (FEBRABAN) [1], in 2021, there was a 23% increase in the volume of transactions through digital channels, which already accounted for 70% of the total.

© 2024 ITMA. All rights reserved.

A report released by International Data Corporation (IDC) [2] indicates that data production doubles every two years, with most of these documents being generated and stored digitally. However, most of the stored files pertain to unstructured data, including images, videos, and raw scanned documents, among others [3].

Despite this shift to a digital environment, it is important to note that many internal processes for identity verification and documentation still use physical documents and images to ensure compliance and mitigate fraud. For instance, the processes for opening checking accounts, even in fintech companies, involve facial biometrics or document scans to verify the identity of clients [4].

In the context of scanned documents discussed in this work, anomalies are any deviations from the standard or expected behavior of the document. They can arise from scanning errors or operational failures, such as blurred or distorted images, missing necessary signatures, or may result from fraudulent attempts to alter the content of the document. Anomaly detection can be useful in ensuring the integrity of documents and trust in business processes, a part of risk management in financial institutions [5] [6].

In the banking sector, the diversity of documents and their constant updates present considerable challenges for identifying different types of possible anomalies. The most popular methods of pattern recognition in documents often require training on large datasets, but the real world does not always allow for such implementations to be protected from failures or even fraud effectively.

The field of computer vision has witnessed significant advances in training models using the concept of Few-Shot Learning (FSL), for example, in facial recognition problems [7], where Artificial Intelligence (AI) increasingly approximates the behavior of natural intelligence, distinguishing and recognizing faces from a few or even a single viewing.

A mechanism with the same advancement, wherein only a few documents are available for Few-Shot Learning (FSL) [8], or even a single document, in a Once Learning [9] framework, to verify a pattern and detect other documents of the same or different patterns, would become the state of the art for anomaly identification tasks in documents. Aiming to advance in this direction, this study explores training methods with a few examples for pattern identification in scanned documents.

© 2024 ITMA. All rights reserved.

Beyond this introduction, the article presents four more sections. The second section provides a review of the state of the art in anomaly detection and pattern identification in documents. The third section describes the proposed work methodology and data used. Subsequently, a study of document classification is presented, and finally, the final considerations and future studies are discussed.

2. Literature Review

Hilal et al. [6] reviewed anomaly detection methods to identify financial fraud. The authors discuss the main techniques used, particularly machine learning techniques, a domain of Artificial Intelligence (AI) in which computational systems learn from data. As mentioned, these studies rely on large datasets to perform this task. Other research areas have explored few-shot methods for anomaly detection tasks.

For some time, the goal of identifying patterns with few examples has been the subject of academic research. In 1999, Weingang [9] proposed the Once-Learning method applied to the training of the Self Organizing Map (SOM) in parallel computing. This method was designed to mimic the way a human brain learns, incorporating elements of quantum mechanics. Li Fei Fei et al., in 2006 [10], used a method based on conditional probability to learn new object categories in computer vision, using the accumulated learning from other categories and called it One Shot Learning.

Wang et al. [8] define Few-Shot Learning (FSL) as a type of machine learning problem where there is a limited number of examples to identify the pattern. The authors also list three typical scenarios for the use of FSL:

- **Learning like a human:** Artificial intelligence seeks to approximate the way humans learn.
- **Learning for rare cases:** Anomalies in documents are often seen as rare cases.
- **Reducing computational effort and cost:** The demand for training models of complex patterns for each case requires processing a large volume of data.

A recent study by Song et al [11] provided a detailed examination of FSL. In this work, the authors segment FSL tasks into Data-Augmentation, which calculates the actual data distribution from examples; Transfer Learning, where learning is transferred from one

© 2024 ITMA. All rights reserved.

dataset to another; Meta Learning, which deals with the way AI can learn, including the possibility of learning new classes never seen before by the AI; and Multimodal Learning, where more than one type of data is used for learning, for example, text and image. This characterization will be detailed in the next section, where studies on document pattern with FSL will be detailed.

For the review of the state of the art, the Scopus knowledge base was used. Searches were conducted for works relating FSL and document pattern identification or anomaly detection in documents or images using the search query "few shot learning" AND ("document pattern" OR (("Document" OR "image") AND "anomaly")). The search returned a total of 30 works. The search results included 19 publications from conferences (10), articles (8), and book chapters (1) related to this research, published from 2020 to the date of the search. Three publications were not accessible and were disregarded in this work. Thus, a total of 16 works on the topic were categorized, as shown in Table 1. Although the studies are not about documents, the task applied in images can be generalized to scanned documents.

Table 1. Published Studies

Topic	Number of Published Studies
Data Augmentation	1
Meta Learning	6
Transfer Learning	8
Multimodal Fusion Learning	1
Total	16

The studies highlight the application of anomaly detection techniques using Few-Shot Learning (FSL) across various research fields, such as healthcare and geosciences. The works will be categorized according to the FSL categories outlined in [11], and further details on the methods and outcomes will be provided subsequently.

2.1. Data-Augmentation

Data augmentation is a method to increase the quantity and diversity of data available for few-shot learning by simulating or generating additional data based on manual rules or machine learning. The goal is to assess the actual data distribution with the highest possible accuracy, thus reducing bias and model overfitting [8].

© 2024 ITMA. All rights reserved.

The only paper addressing this topic is the work by Wismüller et al. [12] which presents an innovative approach for anomaly detection in magnetic resonance imaging (MRI) using self-supervised learning. Instead of relying on labeled data, the proposed method introduces manipulations into the training set by randomly altering the image contrast and removing parts of the input image. This results in an expanded training set, which is then fed into an autoencoder (AE). The AE learns to reconstruct input images, even when presented with varying contrasts and missing parts. This technique enables the learning of relevant image features with few samples, forcing neural networks to leverage contextual information in a model useful for detecting anomalies in medical images, potentially enhancing the productivity of radiologists.

2.2. Transfer Learning

Transfer learning is defined as a method of leveraging prior knowledge learned in a source domain to solve a problem in a related, yet distinct, target domain. The objective is to build mappings from data to labels for specific problems, utilizing pre-training and fine-tuning.

The paper by Dong et al. [13] addresses open-world semantic segmentation, a challenging task that requires segmenting anomalous regions in images and incremental learning for out-of-distribution objects. They propose Region-Aware Metric Learning (RAML) to enhance the integrity of the segmented anomalous regions and Meta-Channel Aggregation (MCA) to further isolate these regions, improving model performance. RAML achieves state-of-the-art in open-world segmentation [13].

Wen et al. [14] propose an anomaly detection method for inspecting steel surface defects in coils, utilizing multi-scale knowledge and a block information module, achieving high performance in detecting anomalies in steel images.

The article by Bijoy et al. [15] presents a FSL model trained with supervised contrastive learning for automated dataset cleansing of image sets. The model distinguishes accurate data from noisy data, enhancing classification performance. The model outperforms other anomaly detection methods.

Huang et al. [16] propose 'Registration Based Few-Shot Anomaly Detection' for anomaly detection with few samples. The method trains a single generalizable model that does not require re-training for new categories. During testing, anomalies are identified by

© 2024 ITMA. All rights reserved.

comparing registered features of the test image with normal support images. The method surpasses few-shot anomaly detection methods in benchmarks.

Tian et al. [17] propose a few-shot anomaly detection method for polyp frames in colonoscopies. The method utilizes few-shot learning and a score inference network to detect anomalies in medical images. The model achieves cutting-edge detection results on a polyp detection dataset, demonstrating its effectiveness.

Wu et al. [18] propose a few-shot learning approach for defect detection in 3D lithium batteries. They utilize a structured light technique based on multiple exposures to reconstruct the 3D shape of the lithium battery and then transform the anomalous parts into 2D images for detection. The method shows good accuracy in classifying surface defects.

2.3. Meta-Learning

Meta-learning is a method of learning to learn, that is, learning an algorithm or model that can quickly adapt to new tasks with few data inputs. The goal is to derive mappings from tasks to target models from specific problems, using episodic training mechanisms. Meta-learning is a general approach that is independent of specific problems, but it depends on the network structure and the similarity between the training and test tasks [19].

Quellec et al. [20] propose a Few-Shot Learning (FSL) framework that extends Convolutional Neural Networks (CNNs) trained for common conditions with an unsupervised probabilistic model for detecting rare conditions in fundus photography. This framework outperforms other approaches, including multitask learning, transfer learning, and Siamese networks, in detecting rare conditions, promoting the adoption of automated screening for eye pathology.

Hansen et al. [21] address the problem of few-shot medical image segmentation using self-supervision and inspiration from anomaly detection. They propose an approach that focuses on segmenting the object of interest, avoiding explicit background modeling, using a single prototype for the foreground class, and calculating anomaly scores for all query pixels. Their model surpasses previous approaches in tasks of segmenting abdominal and cardiac organs in magnetic resonance images with promising results.

© 2024 ITMA. All rights reserved.

Frikha et al. [22] explore the problem of few-shot one-class classification using meta-learning. They adapt the MAML meta-learning algorithm to initialize models suitable for few-shot classification tasks. Their approach outperforms other meta-learning methods, achieving impressive results in various few-shot one-class classification tasks, including anomaly detection in industrial sensor readings.

Baek and Chung [23] propose an object detection model based on the Swin Transformer using explainable meta-learning to minimize fire damage. They combine the Swin Transformer and the YOLOv3 model, applying meta-learning to construct an explainable object detection model.

Huyan et al. [24] present the AUD-Net, a unified deep detector for anomaly detection in hyperspectral images (HSIs) through few-shot learning. Their method focuses on learning relationships between pixels in dual windows and uses a transformer network to estimate relationship embeddings. Experimental results show that the AUD-Net has strong generalization capability across various HSIs and outperforms detectors specifically trained for each HSI.

Jiayao et al. [25] introduce few-shot anomaly detection (FSAD) as a practical anomaly detection method, which provides a limited number of normal images for each class during training. They apply FSAD to training a scalable object detection model without the need for training new classes or tuning parameters, achieving significant improvements in detecting plant diseases compared to existing methods.

2.4. Multimodal Fusion Learning

Multimodal Fusion Learning is a method of integrating information from multiple modalities, such as language, image, and audio, to facilitate few-shot learning. The aim is to overcome the shortcomings of unimodal information, which is limited and insufficient for learning effective features. Multimodal fusion learning can explore semantic space and complement information across modalities [11].

The article by M. Khoshboresh-Masouleh and R. Shah-Hosseini [26] presents an innovative multimodal target detection system based on FSL for the analysis of time-series image data. The challenge lies in interpreting multimodal data, mapping targets and anomalies in time-series images from drone platforms and quadruped robots. Due to the scarcity of training data in these scenarios, the proposed methods need to be

© 2024 ITMA. All rights reserved.

computationally efficient and precise, taking uncertainty measures into account. To address this issue, the authors propose a few-shot learning architecture based on a 'squeeze-and-attention' framework, optimized for the analysis of multimodal time-series images with a small training set. Validated results on three datasets with multiple modalities demonstrate competitive performance, highlighting the effectiveness of this approach in detecting targets in complex remote sensing and photogrammetry scenarios.

Regarding scanned documents, RVL-CDIP [27] is a popular database for document classification studies containing 16 distinct classes of labeled documents. Recent works with good results in document classification frequently use multimodal methods, combining text and image from the document. Wang et al., from JP Morgan AI Research [28], introduced an application of Large Language Models (LLMs) for document pattern identification with a multimodal architecture, following a similar approach to the cited studies that achieved good results with this method. It is noteworthy that in these studies, where there was no interest in FSL, large datasets were used to fine-tune the classification task. This work uses the same database for the results and focuses on the multimodal approach, but with special attention to FSL, following the methodological steps presented in the next section.

3. Methodology

This research is exploratory and quantitative in nature, aimed at assessing the potential of Few-Shot Learning (FSL) methods as a solution to the problem of limited labeled data in anomaly detection. The tests conducted sought to investigate the capability to classify documents in a simulated environment of absent labeled data. Utilizing a widely used dataset that also has applications with large data sets is important to evaluate the method's performance relative to conventional training.

The data used were from a sample of the RVL-CDIP provided by Mohammad [29], consisting of documents from different categories. For this study, the classification of 'invoices' compared to other documents was considered. This explored context simulates a scenario in which a financial institution might expect such a document to be attached to a process and need to identify anomalous documents among them.

For this purpose, the following steps were followed:

© 2024 ITMA. All rights reserved.

- **Two document classes were considered:** Invoices and others (three types of documents were selected for training);

- **Two distinct test samples were selected:** Test 1 with 200 documents (100 from each class) of the same types seen in training; Test 2 with 200 images of document types not seen in training;

- Models were trained using combined information from image and text via OCR (Optical Character Recognition) with pytesseract [30], and extracting the BERT [31] embedding also using the general sample;

- The performance in Test 1 and Test 2 samples of different architectures was compared, always with the availability of 10 normal or out-of-pattern documents.

Given that the sample selection can impact the results obtained, after selecting the model architecture, 30 resampling procedures were conducted, and the results of the performance indicators were consolidated.

4. Results

The initial data processing step applied was the extraction of information from the numerical matrices of the document images. Even with a limited data set, obtaining information from texts is possible using pre-trained models with embeddings embedded in neural networks, similar to the study by Enamoto [32].

For the images, the goal was to approximate the shapes that are perceivable when looking at a document. For this purpose, shapes in quadrants of the image were considered. Figure 1 exemplifies the extraction of shapes from one of the images.

Using these sets of extracted matrices for text and image, architectures were explored to classify documents in the test sample. Thus, results for four different neural network models will be presented, namely: Image Shapes, trained with shapes extracted from the document texts; Pre-Trained Embedding, where a pre-trained model was applied to assign values to the texts; Multimodal, which used data from text and image shapes for training the neural networks; Custom Pre-Trained, which involved pre-training with 500 documents from other categories not known in the training of 'invoices' and subsequently fine-tuned for their detection.

© 2024 ITMA. All rights reserved.

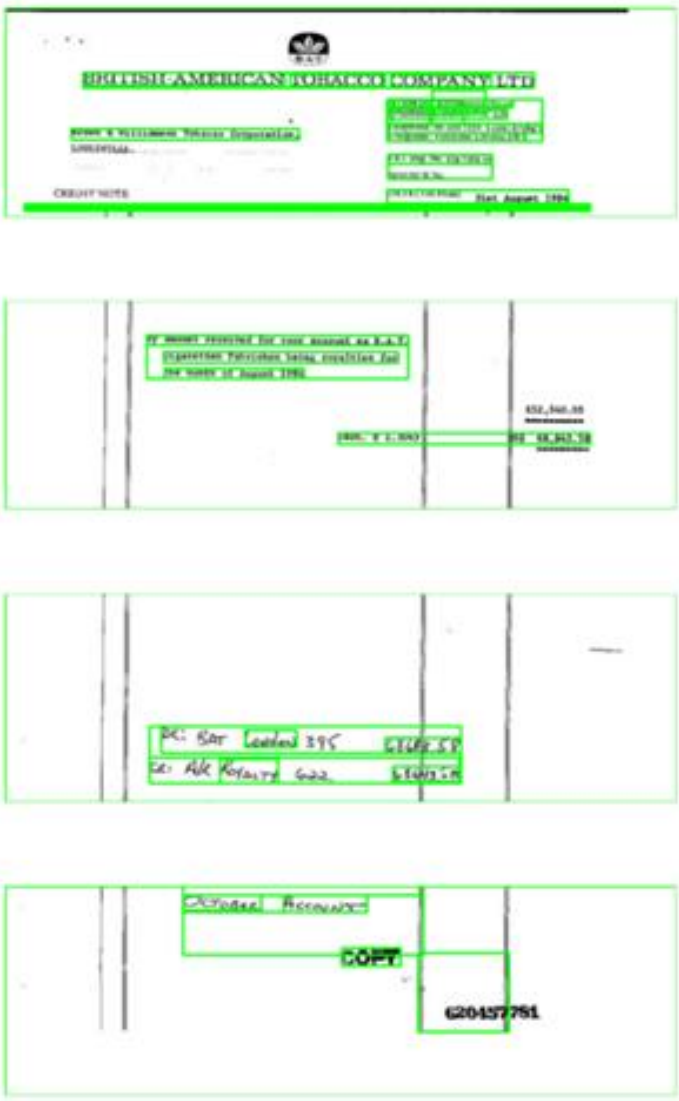


Figure 1. Image Element

The results are compared with the benchmark, constructed with a sample of 30 documents from each class using Google Document AI (GDA) [27]. This benchmark was not constructed as a model comparable to the others, but merely as a result from a well-accepted industry tool. For the models developed in this study, 30 iterations with resampling of 10 documents for training were conducted. Table III presents the median indicators obtained by each method.

© 2024 ITMA. All rights reserved.

Table 2. Model Results

Model	Accuracy	Precision	Recall	F1-Score
Benchmark[1]	0.7010	0.6471	0.8800	0.7457
Simple Image Shapes	0.5817	0.6501	0.4493	0.4880
Multimodal	0.6490	0.6464	0.7107	0.6599
Custom pre-trained	0.6640	0.6709	0.6807	0.6629
Pre-Trained Embeddings	0.7063	0.6811	0.8140	0.7313

[1] Calculated with GDA once with 30 documents from each class for training.

In Table 2, it is observed that the models performed comparably to the benchmark in some cases, even though they were trained with a reduced sample. Figure 2 presents the distributions of the model's accuracy rate, including the rates on Test Sample 2, which is composed of document types not known during the model training. The reference line displays the benchmark result.

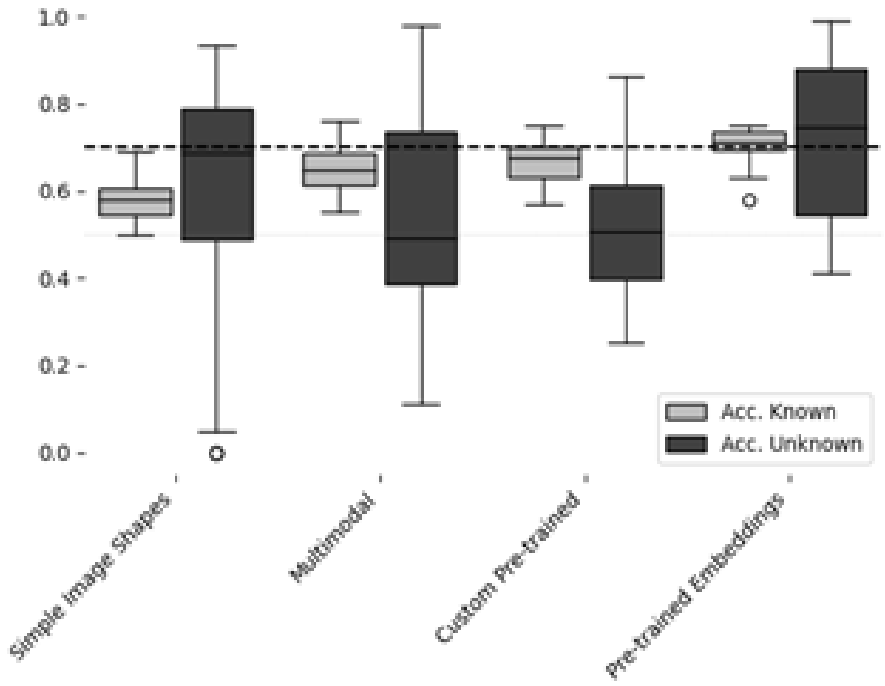


Figure 2. Resampling Results

Despite the lower availability of data, the trained models showed performance in the classification task like these reference lines. It is noteworthy that the text matrix applied

© 2024 ITMA. All rights reserved.

shows the most significant improvement in pattern identification, although the use of the principal shapes of the images also enhances document pattern identification. This result is significant considering the rapid advancement of Natural Language Processing models observed, which can further enhance the transfer of knowledge from other contexts to pattern identification in documents.

An important consideration is that documents in the RVL-CDIP may have a diversity of patterns even within the same category, which could be mitigated in certain applications where the document pattern is more well-defined, a subject that will be evaluated in future studies.

4. Conclusions and Future Studies

In the tasks of anomaly detection in images, the use of Few-Shot Learning (FSL) has shown a strong presence in research in the fields of healthcare, geosciences, among others. This work sought to explore the capability of pattern identification in documents when the dataset is limited.

It was expected that the results of FSL would not reach the levels achieved by studies utilizing large datasets and significant computational processing. However, even with a limited sample, it is evident that the transfer of learning from natural language models and the combination of information in a multimodal model can offer an interesting cost-benefit relationship depending on the application.

A series of improvements can be implemented in this work, such as exploring the performance of more controlled documents, with better-defined patterns like specific contracts. Another aspect to be enhanced in future studies is the more comprehensive exploration of types of anomalies, as this work only considered the identification of documents different from expected, but the capability to identify anomalies such as the absence of specific information, undesirable quality, among others, should also be evaluated.

Future studies intend to explore the scope in which a reduction of the training sample down to the construction of a Once Learning mechanism for anomaly detection in documents is feasible.

Conflicts of Interest: The authors declare no conflict of interest.

© 2024 ITMA. All rights reserved.

References

[1] Febraban. Transações bancárias por canais digitais crescem 23% em 2021 e já são 7 em cada 10 operações no país, Febrabantech. Available online: https://febrabantech.febraban.org.br/temas/inovacao/transacoes-bancarias-por-canais-digitais-crescem-23-em-2021-e-ja-sao-7-em-cada-10-operacoes-no-pais (Accessed on 08 May 2023).

[2] IDC. Data Age 2025: The Evolution of Data to Life-Critical. IDC White Paper, 2018.

[3] RNP. Dados estruturados, não estruturados e semiestruturados, Escola Superior de Redes. Available Online: https://esr.rnp.br/ciencia-de-dados/dados-estruturados-nao-estruturados-e-semiestruturados/ (Accessed on 8 May 2023).

[4] OneSpan. Verificação da identidade digital ajuda a reduzir o atrito e o abandono na abertura de conta. Available online: https://www.onespan.com/pt-br/blog/verificacao-da-identidade-digital-ajuda-reduzir-o-atrito-e-o-abandono-na-abertura-de-conta (Accessed on 08 May 2023).

[5] Yıldız, K.; Dedebek, S.; Okay, F.Y.;Şimşek, M.U. Anomaly Detection in Financial Data using Deep Learning: A Comparative Analysis. Proceedings of the Innovations in Intelligent Systems and Applications Conference (ASYU), Antalya, Turkey, 2022, pp. 1-6.

[6] Waleed Hilal, S.; Gadsden, A.; Yawney, J, Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, **2022**, *Volume 193*, 116429.

[7] Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the gap to human-level performance in face verification, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, USA, 23-28 Jun 2014, IEEE, pp. 1701-1708.

[8] Wang, Y.; Yao, Q.; Kwok, J.T.; Lionel M. Generalizing from a Few Examples: A Survey on Few Shot Learning. ACM Comput. Surv., Vol. 1, No. 1, Article 1, March 2020. Hong Kong University of Science and Technology, Baidu Research, 4Paradigm Inc.

[9] Weigang, L.; Silva, N.C. “A study of parallel neural networks,” in IJCNN’99. International Joint Conference on Neural Networks, pp. 1113-1116, 1999.

© 2024 ITMA. All rights reserved.

[10] Fei-Fei, L.; Fergus R.; Perona P. One-shot learning of object categories, IEEE Transactions on Pattern Analysis and Machine Intelligence , *Volume* 28, no. 4, pp. 594-611, 2006.

[11] Song, Y.; Wang, T.; Mondal, S. K.; Sahoo, J.P. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. arXiv:2205.06743 [cs.LG].

[12] Wismüller, A.; Vosoughi, M.A. Self-Supervised Learning in Medical Imaging: Anomaly Detection in MRI Using Autoencoders, Proceedings of SPIE - The International Society for Optical Engineering, 2022, *Volume* 12102, art. no. 1210206

[13] Dong, H.; Chen, Z.; Yuan, M.; Xie, Y.; Zhao J.; Yu, F.; Dong, B.; Zhang, L. Region-Aware Metric Learning for Open World Semantic Segmentation via Meta-Channel Aggregation. IJCAI International Joint Conference on Artificial Intelligence, 2022.

[14] Wen, X.; Zhao, W.; Yu, Z.; Zhao, J.; Song, K. A Novel Anomaly Detection Method for Strip Steel Based on Multi-Scale Knowledge Distillation and Feature Information Banks Network. Coatings 2023, 13, 1171. Avaiable online: https://doi.org/10.3390/coatings13071171 (Acessed on 10 December 2023).

[15] Bijoy, M.B.; Pebbeti, B.P.; Manoj, A.S.; Fathaah, S.A.; Raut, A.; Pournami, P.N.; Jayaraj, P.B. Deep Cleaner—A Few Shot Image Dataset Cleaner Using Supervised Contrastive Learning. IEEE Access 2023, 11, 18727–18738.

[16] Huang C.; Guan, H.; Jiang, A.; Zhang, Y. Registration Based Few-Shot Anomaly Detection. Lecture Notes in Computer Science, Computer Vision – ECCV, 2022, pp.303-319

[17] Tian, Y.; Maicas, G.; Pu, L.Z.C.T.; Singh, R..; Verjans, J.W.; Carneiro, G. Few-Shot Anomaly Detection for Polyp Frames from Colonoscopy. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020 - 23rd International Conference, pp. 274-284.

[18] Wu, K.; Tan, J.; Li, J.; Liu, C. Few-shot learning approach for 3D defect detection in lithium battery. J. Phys. Conf. Ser. IOP Publ. 2021, 1884, 012024.

[19] Alsadhan, N.A. Value-Added Tax Fraud Detection and Anomaly Feature Selection Using Sectorial Autoencoders, Proceedings of Data Analytics and Management, A. Khanna, Z. Polkowski and O. Castillo, Eds. Singapore: Springer, 2023.

© 2024 ITMA. All rights reserved.

[20] Quellec, G.; Lamard, M.; Conze, P.H; Massin, P.; Cochener, B. Automatic detection of rare pathologies in fundus photographs using few-shot learning, *Medical Image Analysis*, **2020**, *Volume* 61, art. no. 101660, 2020.

[21] Hansen, S.; Gautam, S.; Jenssen, R.; Kampffmeyer, N. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels, *Medical Image Analysis*, **2022**, *Volume* 78, art. no. 102385, 2022.

[22] Frikha, A.; Krompaß, D.; Köpken, H.G.; Tresp, V. Few-Shot One-Class Classification via Meta-Learning, in 35th AAAI Conference on Artificial Intelligence, AAAI 2021, vol. 8B, pp. 7448-7456.

[23] Baek, J.-W.; Chung, K. Swin Transformer-Based Object Detection Model Using Explainable Meta-Learning Mining, in Applied Sciences (Switzerland), vol. 13, no. 5, art. no. 3213, 2023.

[24] Huyan, N.; Zhang, X.; Quan, D.; Chanussot, J.; Jiao, L. AUD-Net: A Unified Deep Detector for Multiple Hyperspectral Image Anomaly Detection via Relation and Few-Shot Learning, in IEEE Transactions on Neural Networks and Learning Systems, pp. 1-15, 2022.

[25] Jiayao, J.; Linfeng, L.; Yunsheng, W.; MingMing, A.; Wenfei, J.; Shipu, X. Cross-class pest and disease vegetation detection based on small sample registration, in IET Image Processing, vol. 17, no. 8, pp. 2299-2308, 2023.

[26] Khoshboresh-Masouleh, M.; Shah-Hosseini, R. Multimodal Few-Shot Target Detection Based on Uncertainty Analysis in Time-Series Images, in Drones, vol. 7, no. 2, art. no. 66, 2023.

[27] Google Cloud. Document AI. Available Oonline: https://cloud.google.com/document-ai (Acessed on 20 November 2021).

[28] Wang, D.; Raman, N.; Sibue, M.; Ma, Z.; Babkin, P.; Kaur, S.; Pei, Y.; Nourbakhsh, A.; Liu, X. DocLLM: A Layout-Aware Generative Language Model for Multimodal Document Understanding. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8529–8548, August 11-16, 2024. Association for Computational Linguistics, 2024.

© 2024 ITMA. All rights reserved.

[29] Mohamed, S. Real World Documents Collections, Version 1. Available online: https://www.kaggle.com/datasets/shaz13/real-world-documents-collections?resource=download (Acessed on 10 July 2023).

[30] Google. Pytesseract: Python-tesseract is an optical character recognition (OCR) tool for Python. GitHub, 2023. https://github.com/madmaze/pytesseract. (Acessed on 10 December 2023).

[31] Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.https://arxiv.org/abs/1810.04805. (Accessed on 10 December 2023).

[32] Enamoto, L., Weigang, L., Filho, G. P. R. (2021). Generic framework for multilingual short text categorization using convolutional neural network. Multimedia Tools and Applications, 80(9), 13475–13490.

© 2024 ITMA. All rights reserved.