ANOMALY DETECTION VIA CONTEXT AND LOCAL FEATURE MATCHING

Antanas Kascenas[1,2] *Rory Young*[2]
Bjørn Sand Jensen[2] *Nicolas Pugeault*[2] *Alison Q. O'Neil*[1,3]

[1] Canon Medical Research Europe, Edinburgh, UK
[2] University of Glasgow, Glasgow, UK
[3] University of Edinburgh, Edinburgh, UK

ABSTRACT

Unsupervised anomaly detection in medical imaging is an exciting prospect due to the option of training only on healthy data, without the need for expensive segmentation annotations of many possible variations of outliers. Most current methods rely on image reconstruction error to produce anomaly scores, which favors detection of intensity outliers. We instead propose a discriminative method based on a deep learning self-supervised pixel-level classification task. We model context and local image feature information separately and set up a pixel-level classification task to discriminate between positive (matching) and negative (mismatching) context and local feature pairs. Negative matches are created using data transformations and context/local shuffling. At test-time, the model then perceives local regions containing anomalies to be negative matches. We evaluate our method on a surrogate task of tumor segmentation in brain MRI data and show significant performance improvements over baselines.

Index Terms— Anomaly detection, Unsupervised learning, Self-supervised learning, MRI, Deep learning.

1. INTRODUCTION

In this paper we consider the problem of anomaly detection, specifically the detection and localization of focal pathologies. Automated pathology detection could play a valuable role in computer aided diagnosis (CAD) e.g. for scan triage, as a second read, and for identifying incidental findings. However, it is challenging to acquire a comprehensive training dataset containing examples of every possible appearance of every possible pathology. Unsupervised anomaly detection (UAD) methods learn only from healthy subject data, with the goal of detecting anomalies (outliers) at test time.

Existing UAD methods in medical imaging include methods based on classification [1], restoration [2], [3], generative adversarial networks (GANs) [4]–[7], and reconstruction error based autoencoders (AEs) [8]–[16]. Most existing methods make use of image reconstruction error to produce pixel-level anomaly scores at test time, relying on the assumption that anomalous regions are going to be reconstructed more poorly than healthy regions. Such methods might be limited to anomalies presenting significant pixel intensity deviations [17]. Additionally, it has been found that, in practice, reconstruction error based models can generalize and sometimes reconstruct even unseen anomalous regions with little error. Thus, non-reconstruction approaches might be needed to tackle the detection of harder anomalies. Discriminative models for anomaly detection have already shown some success both in computer vision [18] and medical imaging [1].

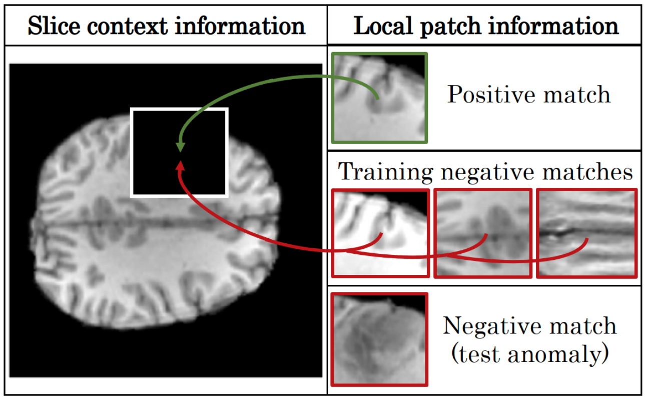

Fig. 1. Context and local feature matching. The method is trained to discriminate positive and negative pairs of context and local features. Training negative pairs are generated using intensity/spatial transformations and shuffling.

In this paper, we propose an anomaly detection method based on a novel self-supervised pixel-level task, *context to local feature matching (CLFM)*, of learning to match pairs of context and local image feature information using healthy data (see Fig. 1). Our contributions are as follows:

- We propose the novel self-supervised CLFM task of predicting matches of context and local information in medical images, which enables a discriminative modeling approach to anomaly detection and localization.
- We compare our proposed method to state-of-the-art anomaly detection methods on MRI brain tumor data and achieve superior unsupervised generalization.

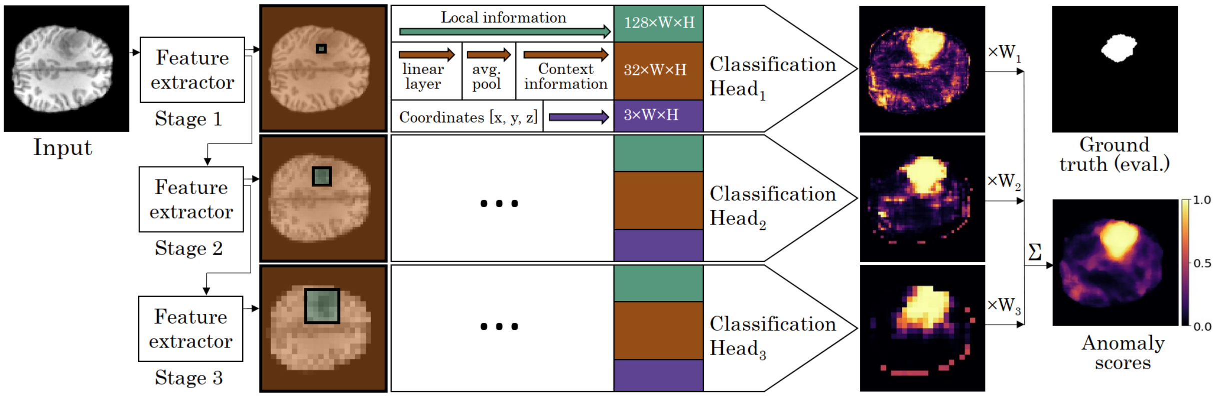

Fig. 2. Hierarchical method configuration of the CLFM approach. Convolutional feature extractors and classification heads operate at three scales. Scores from each stage are bilinearly upsampled and combined via a weighted mean.

2. METHOD

Our approach to UAD is based on separation of local (i.e. local neighborhood) and context (i.e. surrounding image) information. We enforce exclusivity of information between local and context features by leaving a buffer between the two regions that ensures contiguous and non-overlapping receptive fields between their convolutional representations. This exclusivity is required to prevent trivial solutions to the self-supervised context and local information matching. We then train on the self-supervised CLFM classification task, requiring the model to learn the matched (i.e. healthy) pairings of local and context information. In the absence of real anomaly training examples, we synthesize mismatched (i.e. anomalous) pairs. Finally, to present the appropriate balance of local and context information for a wide range of anomalies, we use a hierarchical approach where we adjust the receptive field of local information associated with each pixel. We describe each part of the system below.

2.1. Local and context feature extraction

We apply a shallow CNN to learn the **local features** corresponding to each pixel in the image. The **context features** are constructed by aggregating the local information across the context region i.e. the whole image excluding the local region, with a buffer that prevents receptive field overlap. We perform the aggregation by linearly projecting the local features and averaging over the context region.

The requirement for exclusivity between local and context information prevents us from using standard neural network normalization methods such as batch or layer normalization, which normalize across the whole image. Instead, we use a combination of weight standardization and L_2 normalization across the channel dimension.

2.2. Negative pair generation

For generating negative pairs, we employ a few strategies:

1. Shuffle the patches (i.e. extracted features) across each training image batch to give out-of-context matches.
2. Extract mismatched patches from an image augmented with intensity transformations. We use additive intensity transformations in the range of -0.15–0.15 and multiplicative transformations in the range of -1.3–1.3.
3. Extract mismatched patches from a combination of heavily augmented images randomly selected from the training data. We use intensity transformations, rotations, flips, resizing, cropping and blurring to generate negatives.

2.3. Pair classification

A classification head is trained to output the match probability of the context and local information pair at every pixel. The classification head has 3 concatenated pixelwise inputs: context features, local features, and the x, y, z volume coordinates. The output probabilities p are used for binary cross-entropy loss (BCE) for training and as anomaly scores during inference. The pixelwise loss is calculated using the binary pair labels t (1 for natural pairs in healthy slices, 0 for synthesized negative pairs), averaged over the stage i brain foreground pixels (i.e. non-zero in any modality) F_i and summed over the stages:

$$\text{Loss} = \sum_{i=1}^{3} W_i \frac{1}{|F_i|} \sum^{F_i} \text{BCE}(p, t)$$

We use a positive to negative pair ratio of $1 : 2$ during training.

2.4. Hierarchical configuration

Shallow CNNs with limited receptive fields may struggle to identify larger or more complex anomalies. Thus, we apply our method in a hierarchical configuration using three stages (see Fig. 2). Each stage bilinearly downsamples the local information learned by the CNN of the previous stage and applies a new CNN to learn from an effectively expanded receptive field with respect to the original resolution. At all scales, context features are then computed and the patch is classified. We then combine the classification results from the three stages by bilinearly upsampling all of the results to the original resolution and using a weighted mean where the weight w_i for each stage i is $W_i = 2^{-i}$.

3. EXPERIMENT SETUP

3.1. Dataset

There is a lack of public datasets for evaluating anomaly detection approaches. Therefore, we evaluate on the surrogate task of brain tumor segmentation using data from the BraTS 2021 challenge [19]–[21]. This data comprises native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes for each patient from a variety of institutions and scanners, which has already been co-registered, skull-stripped and interpolated to the same resolution. Labels are provided for tumor sub-regions: the GD-enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor.

We split the dataset into 938 training, 62 validation, and 251 test patients. We consider the union of the tumor labels to be the anomalous regions. During training, we only use slices that do not contain any tumor pixels, under the assumption that they represent healthy tissue. For the data input to the models, we concatenate all four modalities at the channel dimension for each patient. We scale the pixel intensity values in each modality of each scan by dividing by the 99th percentile brain pixel intensity. All slices are downsampled to a resolution of 128×128 (1.62mm/pixel).

3.2. Baselines

We chose three of the best performing UAD methods as evaluated by Baur et al. [22]. Namely, we use f-AnoGAN [4], a GAN-based approach as well as a variational autoencoder (VAE) method evaluated using the standard reconstruction error [9], [11] and restoration [2] methods for producing anomaly scores.

3.3. Implementation details

CLFM model: As described in Section 2, our model comprises three stages, each made up of a CNN, local-to-context projection head, and a classification head. The multi-scale (multi-stage) architecture for learning local information is similar to a standard encoder configuration, with blocks of 2 convolutional layers (the CNNs) connected by bilinear downsampling layers.

More precisely, the feature extractor CNNs comprise two weight standardized [23] convolutional layers with 128 output channels, a kernel size of 3×3, Swish activations and L_2 normalization across the channel dimension. The local-to-context projection heads are convolutional layers of kernel size 1 that project CNN outputs into context averaging space with 32 dimensions. Finally, the classification head uses the same architecture as the previously described CNNs but with kernel sizes of 1 and a final convolutional classification layer of kernel size 1 that projects into a single dimension representing the context and local information match probability. The model is trained using the binary cross entropy loss (see Section 2.3). We train the model using the Adam optimizer with a cosine annealed maximum learning rate of 0.001 and batch size of 16 for 160,000 iterations. We use stochastic weight averaging [24] with a linear annealing schedule converging to a learning rate of 0.0001 in the last 38,400 iterations to produce the final model.

f-AnoGAN: We adapt the original public implementation [1] for the brain MR data task as follows. We use an additional generator, discriminator and encoder block to account for the higher resolution. Strided convolutions and transposed convolutions are used for downsampling and upsampling respectively. We use a batch size of 32 and learning rates of 0.001, 0.001, 0.00001 for the generator, discriminator and encoder respectively. The encoder was trained using $\kappa = 1 \times 10^{-8}$.

VAE: We implement an encoder-decoder architecture with three downsampling/upsampling stages and a bottleneck with dimensionality of 128. Each encoder stage consists of two weight-standardized convolutions [23] with kernel size of 3 and 64, 128, 256 output channels for the three stages respectively followed by Swish activations followed by group normalization layers with groups of 8, 16, 32 respectively. Average 2× pooling is used for downsampling. The decoder architecture mirrors the encoder in reverse, using transposed convolutional layers for upsampling. We use the sum of $L2$ reconstruction error and KL-divergence with a weight of $\beta = 0.001$ as the loss. Training is done using the Adam optimizer with a cosine annealed maximum learning rate of 0.0001, batch size of 16 and train for 160,000 iterations. We use stochastic weight averaging [24] with a linear annealing schedule converging to a learning rate of 0.00001 in the last 38,400 iterations to produce the final model.

VAE restoration: Using the VAE model described above, we implement a restoration method [2] to produce the anomaly scores. We perform the restoration procedure using 100 iterations on individual slices basing our implementation on public source code [2]. Note that due to the iterative nature of the restoration procedure it takes significantly longer (approx. ×100) to produce predictions compared to other methods.

[1] https://github.com/tSchlegl/f-AnoGAN

[2] https://github.com/yousuhang/Unsupervised-Lesion-Detection-via-Image-Restoration-with-a-Normative-Prior

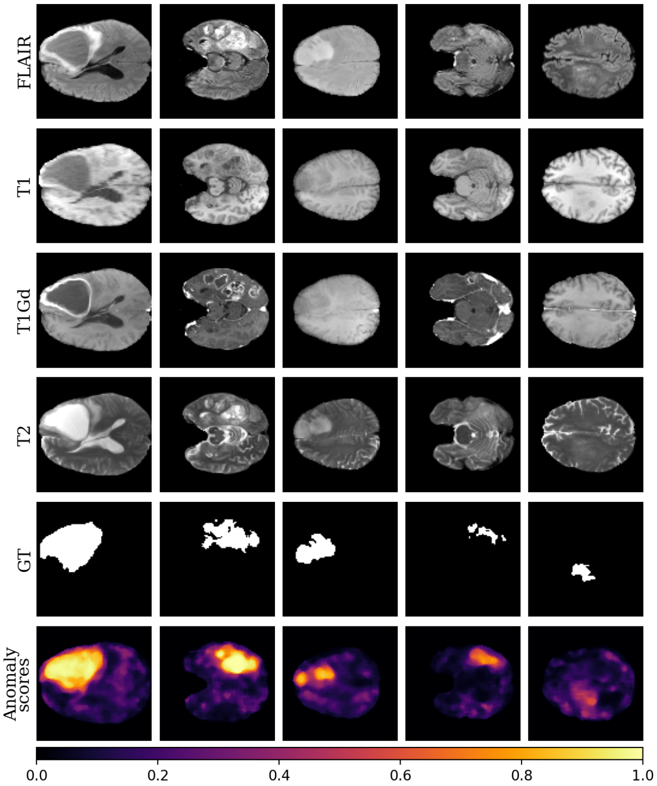

Fig. 3. Sample CLFM model results, (left to right) from easy (obvious) to difficult (indistinct) anomalies.

For all methods, we use median filtering with a kernel size of 5 as a postprocessing step to reduce high frequency noise in the predicted anomaly scores. Slight rotation, brightness, flip and stretch data augmentation during training was found to slightly improve performance of VAE and CLFM methods.

4. RESULTS

We evaluate the **anomaly segmentation accuracy** of our method against the baselines using the area under the precision-recall curve (AUPRC) at the pixel level which allows evaluation without setting an operating point for the produced anomaly scores. We also calculate $\lceil\text{Dice}\rceil$, a Dice score which measures the segmentation quality using the optimal operating point found using the validation set ground truth. We include an ablation study investigating the effects of context information, coordinate inputs and multiple stages in Table 1. Table 1 shows that CLFM outperforms the baselines, with the contextual information and multi-scale components playing an important role in its success. Fig. 3 shows visualizations of the predictions for a range of anomalies.

We further evaluate **anomaly localization accuracy** at the scan level in order to reflect the more realistic scenario where anomalies need to be localized but not necessarily precisely segmented. We use the optimal operating point found using the validation set to binarize the test predictions for each model. We then calculate the Dice scores for each test patient and consider the segmentations above the thresholds of 0.5 or 0.75 as positive localizations. Test scans where tumor segmentations are worse are considered to be false negatives. We thus report the test recall at the patient level in Table 2.

Method	AUPRC	$\lceil\text{Dice}\rceil$
f-AnoGAN [4]	0.365±0.024	0.449±0.014
VAE (recon.) [8], [11]	0.554±0.006	0.538±0.004
VAE (restoration) [2]	0.767±0.002	0.703±0.002
CLFM (ours)	**0.811±0.002**	**0.742±0.001**
CLFM−Ctx.	0.613±0.019	0.613±0.012
CLFM−Coord.	0.731±0.003	0.681±0.001
CLFM−$S_3 - S_2$	0.748±0.003	0.689±0.002
CLFM−S_3	0.800±0.001	0.731±0.001

Table 1. Voxelwise tumor segmentation results as measured by area under the precision-recall curve and optimal Dice score. ± indicates standard deviation across three runs. −Ctx., −Coord., $-S_3 - S_2$ refer to the model without context aggregation, coordinate inputs, and excluding contributions from stage 2 and 3 classification heads respectively.

	Recall	
Method	@0.5$\lceil\text{Dice}\rceil$	@0.75$\lceil\text{Dice}\rceil$
f-AnoGAN [4]	0.36±0.03	0.01±0.00
VAE (recon.) [8], [11]	0.57±0.00	0.08±0.01
VAE (restoration) [2]	**0.91±0.01**	0.38±0.01
CLFM (ours)	**0.90±0.00**	**0.61±0.01**

Table 2. Scan-level results as measured by recall at 0.5$\lceil\text{Dice}\rceil$ and 0.75$\lceil\text{Dice}\rceil$ thresholds for successful localization. ± indicates standard deviation across three runs.

5. CONCLUSION

This work presented a novel self-supervised system for detecting and localizing anomalies in brain MR images based on discriminative modeling of context and local information pairs. We showed that by generating the appropriate negative (mismatched) pairs during training we can obtain a model that is effective at detecting anomalous lesions in the brain. Our method uses no manual dense annotations and only needs healthy data to be trained. Discriminative anomaly detection methods like ours are more aligned with the fields of image segmentation and classifications than more traditional and restrictive autoencoder and reconstruction error based methods. Thus, discriminative methods are easier to integrate with the rapid advances in large-scale self-supervision, pretraining, fine-tuning, and semi-supervision that will lead to more practical anomaly detection applications in the future.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by the BraTS'21 challenge. Ethical approval was not required as confirmed by the license attached with the open access data.

References

[1] J. Tan *et al.*, "Detecting outliers with foreign patch interpolation," *arXiv preprint arXiv:2011.04197*, 2020.

[2] X. Chen *et al.*, "Unsupervised lesion detection via image restoration with a normative prior," *Medical image analysis*, vol. 64, p. 101 713, 2020.

[3] S. N. Marimont *et al.*, "Anomaly detection through latent space restoration using vector-quantized variational autoencoders," *arXiv preprint arXiv:2012.06765*, 2020.

[4] T. Schlegl *et al.*, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Medical image analysis*, vol. 54, pp. 30–44, 2019.

[5] C. Baur *et al.*, "SteGANomaly: Inhibiting CycleGAN steganography for unsupervised anomaly detection in brain MRI," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 718–727.

[6] C. Han *et al.*, "MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction," *BMC bioinformatics*, vol. 22, no. 2, pp. 1–20, 2021.

[7] V. Alex *et al.*, "Generative adversarial networks for brain lesion detection," in *Medical Imaging 2017: Image Processing*, International Society for Optics and Photonics, vol. 10133, 2017, 101330G.

[8] H. E. Atlason *et al.*, "Unsupervised brain lesion segmentation from MRI using a convolutional autoencoder," in *Medical Imaging 2019: Image Processing*, International Society for Optics and Photonics, vol. 10949, 2019, 109491H.

[9] C. Baur *et al.*, "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 161–169.

[10] V. Alex *et al.*, "Semisupervised learning using denoising autoencoders for brain lesion detection and segmentation," *Journal of Medical Imaging*, vol. 4, no. 4, p. 041 311, 2017.

[11] D. Zimmerer *et al.*, "Unsupervised anomaly localization using variational auto-encoders," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 289–297.

[12] D. Zimmerer *et al.*, "Context-encoding variational autoencoder for unsupervised anomaly detection," *arXiv preprint arXiv:1812.05941*, 2018.

[13] X. Chen *et al.*, "Unsupervised detection of lesions in brain MRI using constrained adversarial autoencoders," *arXiv preprint arXiv:1806.04972*, 2018.

[14] N. Pawlowski *et al.*, "Unsupervised lesion detection in brain CT using Bayesian convolutional autoencoders," *Medical Imaging with Deep Learning*, 2018.

[15] C. Baur *et al.*, "Bayesian skip-autoencoders for unsupervised hyperintense anomaly detection in high resolution brain MRI," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2020, pp. 1905–1909.

[16] L. Zhou *et al.*, "Unsupervised anomaly localization using VAE and beta-VAE," *arXiv preprint arXiv:2005.10686*, 2020.

[17] F. Meissen *et al.*, "Challenging current semi-supervised anomaly segmentation methods for brain MRI," *arXiv preprint arXiv:2109.06023*, 2021.

[18] D. Hendrycks *et al.*, "Deep anomaly detection with outlier exposure," *Proceedings of the International Conference on Learning Representations*, 2019.

[19] U. Baid *et al.*, "The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv:2107.02314*, 2021.

[20] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BraTS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[21] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.

[22] C. Baur *et al.*, "Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study," *Medical Image Analysis*, p. 101 952, 2021.

[23] S. Qiao *et al.*, "Weight standardization," *arXiv preprint arXiv:1903.10520*, 2019.

[24] P. Izmailov *et al.*, "Averaging weights leads to wider optima and better generalization," English (US), in *34th Conference on Uncertainty in Artificial Intelligence*, 2018, pp. 876–885.