ARTICLE TEMPLATE

Automatic pancreas anatomical part detection in endoscopic ultrasound videos

Antoine Fleurentin[a] and Jean-Paul Mazellier[a, b] and Adrien Meyer[b] and Julieta Montanelli[a] and Lee Swanstrom[a] and Benoit Gallix[a] and Leonardo Sosa Valencia[a] and Nicolas Padoy[a, b]

[a]IHU Strasbourg, France; [b]ICube, University of Strasbourg, CNRS, France

ARTICLE HISTORY
Compiled September 14, 2022

ABSTRACT
Pancreatic cancer, due to poor survival rates, stands as the 3^{rd} cause of death by cancer in 2021 and is predicted to become 2^{nd} in 10 years from now. To counter this trend, early diagnosis is the only way to significantly improve survival rate but as standard 3D medical imaging fails to achieve early detection, endoscopic ultrasound (EUS) stands as the only viable option today. This technique, however, is not widely available due to its difficulty. One challenging aspect of EUS is the complex interpretation of ultrasound images during the examination. It is not rare for non-experts to miss the screening of parts of the pancreas during the procedure, leaving tumors undetected. Here, we propose an automated method to support non-expert clinicians in their practice by providing a deep learning (DL) based tool able to detect the anatomical parts seen under EUS in order to guide non-expert EUS practitioners. For this purpose, we have collected 41 EUS videos and annotated the anatomy viewed in each video frame. Considering the challenging and novel nature of EUS data and motion, we propose a systemic analysis of state-of-the-art feature extractors and temporal modules on this unique dataset. In particular, we extend popular models with LSTM temporal modules and compare their performance to the newly introduced vision transformers, yielding an overall comparison of 35 models. The results highlight the benefits of transformers for their ability to capture more anatomical context thanks to the division into patches and positional embeddings. As a result, our study paves the way to AI-assisted pancreas examination for early cancer detection.

KEYWORDS
Pancreas ; Ultrasound ; Video classification ; Convolutional neural network ; Vision transformers

1. Introduction

Pancreatic cancer is one of the cancer associated with the poorest prognosis. It has a survival rate at 5 years as low as 8% (Ferlay et al. 2019). However, this dramatic figure is highly dependent on the tumor size at first diagnosis examination. Indeed, identification of tumors smaller than 10mm can increase the survival rate by a factor of 10, up to 80% (Ferlay et al. 2019; Tonozuka et al. 2021b). This stresses the need for an efficient early diagnostic route to maximize patient's treatment efficiency. Unfortu-

CONTACT J.-P. Mazellier. Author. Email: jean-paul.mazellier@ihu-strasbourg.eu

nately, standard imaging methods such as computed tomography scanner (CT-scan) or magnetic resonance imaging (MRI) are not yet able to reveal tumors smaller than 20 mm (Huang et al. 2021). Pancreatic cancer evolution worldwide is today considered as a pandemic issue with a number of new cases doubling every decade since the year 2000 (Huang et al. 2021). Currently, even if pancreatic cancer is the 7^{th} cause of cancer in Europe, it stands as the 3^{rd} cause of cancer-related death and its rapid evolution may bring it to the 2^{nd} rank in 2030's. In this context, endoscopic ultrasound (EUS) currently stands as the only efficient technique for detecting early signs of pancreatic cancer. In particular, it has been demonstrated to be efficient on tumors as small as 5mm (Huang et al. 2021). The main drawback of this method is its important dependence on the clinician's skill. Many years of practice are necessary to fully exploit its diagnosis potential. The difficulty lies in the duality of the technique, as the clinician must (Zhang et al. 2018; Białek and Jakubowski 2017): (1) navigate the flexible probe inside the patient body in a weakly constrained configuration (the stomach), and (2) interpret complex ultrasound images at the same time. For a non-expert endoscopist, identifying parenchyma of a non-normal pancreas in ultrasound images already represents a highly complex task and it is one of the first skills to acquire during EUS training. Today, no automated or semi-automated system is available for such task. On top of pancreas parenchyma recognition, practitioners have to identify which segment of the gland is under consideration. This is mandatory for navigating the EUS probe and, as a result, assessing screening of the complete gland. This coverage assessment is the only way to confirm/deny the presence of lesions or signs of early cancer. Overlooking a part of the pancreas could imply the non-diagnostic of a lesion, which would lead to dramatic consequences for the patient (Adler and Diehl 2015). An automated assisting system for pancreas parenchyma presence detection and segment identification, as could be provided by an artificial intelligence based assistant, would help non-expert EUS practitioners in their daily task.

Verifying the correctness of the examination of an organ is found in other domains. For instance Freedman et al. (2020) highlight the usefulness of a tool that verifies that every part of the colon has been observed to avoid missing polyps. Wu et al. (2017) highlight how DL can be used to ensure the image quality for obstetric examination, which is crucial for accurate biometric measurement. In terms of research on DL model applied to EUS, very few studies have been produced. Tonozuka et al. (2021a) develop a computer-assisted diagnosis system to localise pancreatic ductal carcinoma. Moreover, (Zhang et al. 2020) study multiple chained DL-based models to identify stations and segments the pancreas parenchyma. But none of these methods take advantage of the temporal information from EUS videos, which is of key importance to the practitioner. It is admitted in the endoscopist community that several anatomical structures are almost impossible to identify from a static image. This arise from the difficulty to replace the current frame in a local context where different external visual markers are needed to disambiguate pancreas parts classification. This is in particular the case for differentiating the pancreas body from the tail which is very complicated to distinguish by their own. Nevertheless, vicinity of the spleen or kidney is indicative of tail part. On the other hand, surrounding structures like the common bile duct or mesenteric vessels are also distinctive markers. But as US can only give a 2D cross-sectional view per frame, multiple consecutive frames, corresponding to temporal context, may be required to gather all these information for final classification by human experts.

Our aim is to pave the way for an AI assistant to EUS procedures. We first introduce a new dataset with EUS videos collected from hepato-biliary surgery department ***** and labelled with the examined segment of the pancreas as well as the liver.

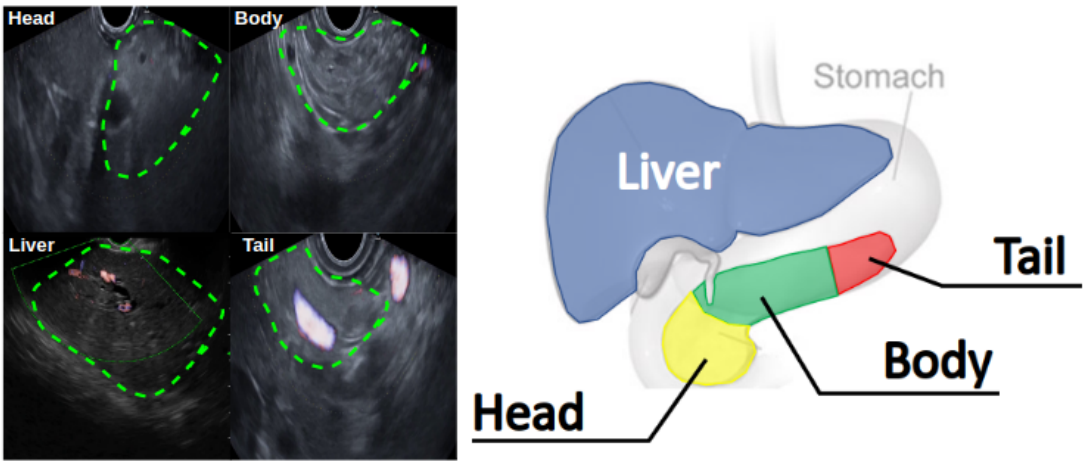

Figure 1. Left: Frames samples illustrating classes *liver*, *head*, *body* and *tail*. In each image is highlighted the corresponding organ approximate boundary (dashed bounded region) for reader's convenience. Right: Diagram of the anatomical representation of liver as well as pancreas head, body and tail segments.

We then propose a systematic analysis of state-of-the-art models and proposed extensions to analyze challenging EUS data. We focus our methodology on the extensive evaluation of 20 state-of-the-art DL architectures (12 based on CNN, 8 based on visual transformers) in order to capture the main properties of spatial context learning. On top of this spatial information, we also studied the temporal component and its influence on EUS clip classification performance. In total, 35 DL models are studied in this work.

2. EUS-C40 Dataset

2.1. Videos collection

We collected clinical videos from 40 patients undergoing EUS for pancreatic cancer screening. These patients previously underwent a pancreas control CT-scan revealing a pancreas disease suspicion, which can be one of the primary marker for possible pancreas cancer. Our video database is therefore constituted with views of pathological pancreas (and also, generally, liver) parenchyma. All videos start when the EUS probe has reached the stomach and end when the EUS probe is removed from the patient's body. The videos length are between 15 and 60 minutes and they come from 2 different ultrasound systems (Olympus and Hitachi EUS systems). The frequency of the ultrasound signal was set at 7.5 MHz for all the EUS. All videos were resized to 224x224 pixels and any layout from ultrasound systems was removed (including potential patient information). We also point out that these videos were recorded from EUS procedures performed by an expert endoscopist. Consequently, videos are relatively smooth in the sense that the practitioner masters the flexible EUS movement control, which allows to have a continuous screening of the pancreas parts (without erratic probe movement that could render in fuzzy EUS videos) and ensures a complete gland screening with high degree of confidence.

2.2. Annotation

An experienced EUS endoscopist annotated EUS video frames with the following labels (see Figure 1):

- Liver parenchyma: the liver has a characteristic parenchyma, different from pancreas' one. It is generally easy to identify by a trained clinician.
- Pancreas head parenchyma: the head can be identified from both its texture and the presence of the common bile duct (as in Figure 1), mesentric blood vessels and/or portal vein.
- Pancreas body parenchyma: when examining the body in EUS video, internal structures such as the Wirsung duct (a bile duct) or splenic vessels are often visible. The kidney (external structure) can sometimes be seen towards the left pancreatic body.
- Pancreas tail parenchyma: like the body, the pancreas tail can be characterized by internal structures such as splenic vessels. But a disambiguation criteria stands in the observation of typical external anatomical landmarks such as adrenal gland and/or spleen.
- Artefacts: Doppler mode obscuring the ultrasound image, large shadow cone due to poor probe position and probe outside the patient.
- Other: Other anatomical structures likely to be encountered during an EUS (gastric folds, stomach wall, mediastinum, duodenum).

The videos are processed at 5 fps (with initial EUS videos collected at 30fps). Ambiguous EUS video parts and portions with elastography were excluded (15% of the initial frames). The annotated data represents 179 092 frames, corresponding to 786 video clips containing the same label continuously. The average clip duration is 37 seconds.

We point out here the role of spatio-temporal context in the annotation process as performed by EUS expert. Indeed, if liver parenchyma is quite distinctive (by its texture and shape) and straightforward to annotate with high confidence, pancreas parts are much more complicated to label. The difficulty lies in its texture that is less characteristic with respect to the rest of EUS images, but also the distinction between the different parts (*head*/*body*/*tail*). The volumes of pancreas segments are not equal and are distributed in about (1) head: 40%, (2) body: 40% and (3) tail: 20%. This typical distribution stands for a normal gland, but in case of ill pancreas, as in our data, it frequently appears that the tail is atrophic and its relative volume is lowered. This translates in our dataset into almost 5 times less *tail* segments with respect to *head* and *body* segments. *head*, *body*, *artefact* and *other* segments are otherwise balanced (cf Figure 2). Beyond *tail* occurrence imbalance with respect to other classes, the typical parenchyma size is also smaller in this region compared to rest of the pancreas parenchyma. This translates to a smaller surface area of tail parenchyma in EUS videos compared to head, body or even liver. Eventually, regarding *tail* classification, it turns out that its determination is highly dependent on external anatomical landmarks in order to be distinguished from the body part.

3. Method

Based on our experience in annotating pancreas segments, we aim to demonstrate that temporal information is highly valuable in EUS and show the challenge of ob-

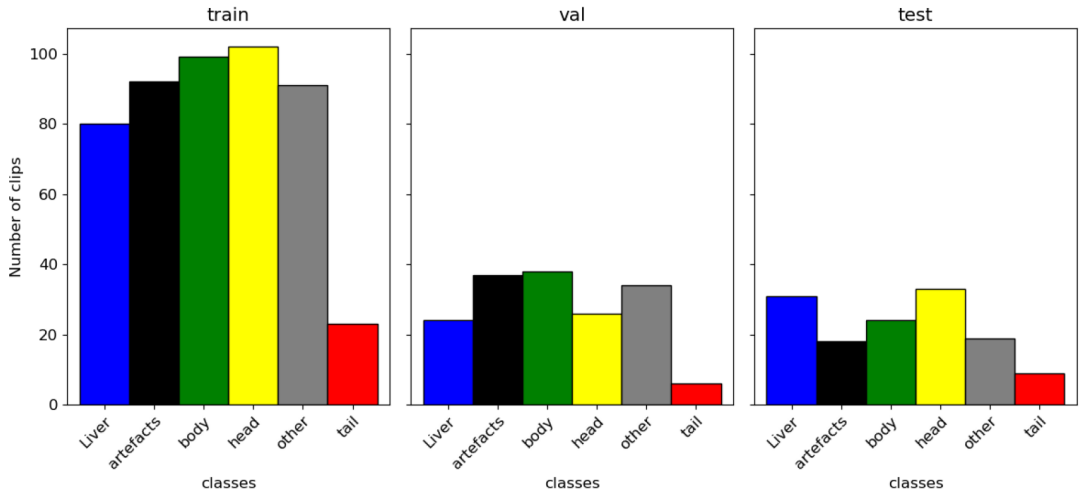

Figure 2. Classes distribution for the different datasets.

taining exploitable spatial features from ultrasound images. To better understand DL models ability to extract these spatial features, we have chosen to compare CNN and vision transformer (ViT) architectures. CNN allowed the renewal of DL in computer vision, with associated increase in computing power and availability of large datasets (Krizhevsky et al. 2017). CNN in particular benefit from the local pixel correlation analysis and are known to offer shift invariance; nevertheless, they suffer from long-range correlation impediment which has been treated by combining multi-scale analysis (Redmon and Farhadi 2018)(Lin et al. 2017). Aside CNN, computer vision community has recently focused on new ViT architectures in which the paradigm of local pixel correlation, treated by convolutions, has been replaced by self-attention maps approach with full-image range correlation analysis (Dosovitskiy et al. 2020)(Carion et al. 2020). These new ViT architectures demonstrate significant improvement in many computer vision tasks. However, both CNN and ViT performances has been mainly demonstrated on natural images, which significantly differ from the ultrasound images we are interested in. Furthermore, this comparison is motivated by the intrinsic behavior of ViT. Indeed, we made the hypothesis that patching and position embedding will better capture connections between large textures and rare areas with salient objects in the ultrasound images. For our work, three CNN (Tan and Le 2019)(He et al. 2016)(Liu et al. 2022) and five ViT (Touvron et al. 2021)(d'Ascoli et al. 2021)(Chu et al. 2021)(Ali et al. 2021)(Bao et al. 2021) architectures have been selected. In total, we have evaluated 15 models as we considered several *flavors* of the same backbone (cf Figure 3-C1), allowing us to test for the influence of the number of parameters on classification efficiency. Concerning the temporal features, we first tried to dissociate them from the spatial ones. For that purpose, starting from the very same backbones with created 15 new models incorporating a Long Short-term Memory (LSTM) (Hochreiter and Schmidhuber 1997) to handle the temporal component in addition to spatial features. Here the LSTM have been used to aggregate feature, as outputed from CNN backbones, through time in order to proceed an internal representation used for final classification (cf Figure 3-C2). Then we went further by combining these two aspects with other models designed for video classification. For this type of classifier, we selected four models based on CNN (Karpathy et al. 2014)(Tran et al. 2018)(Feichtenhofer et al. 2019)(Feichtenhofer 2020) and one model based on ViT (Fan et al.

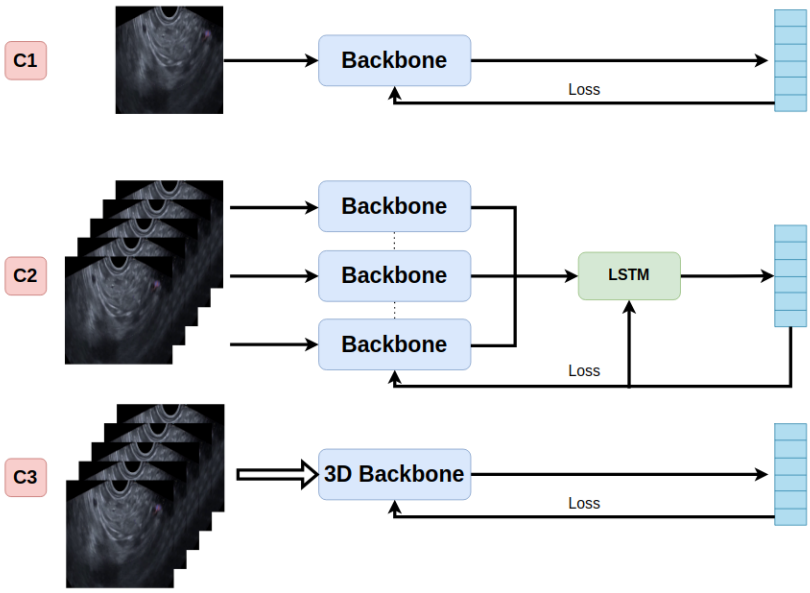

Figure 3. Diagram illustrating the different configurations.

2021) (cf Figure 3-C3). This study thus integrates 35 models in total and provides their extensive comparison in classifying EUS video clips.

In order to take advantage of the temporal aspect in our video data, subclips were used to train and evaluate DL models. A subclip, randomly selected from a clip, are constituted from 16 consecutive frames from downsampled EUS video at 5fps. This corresponds to 3.2 seconds of real-time EUS video. In our experiments, the 3 tested configurations can be described as follow (see Figure 3 for representation of training phase):

- **Configuration 1** (C1): A classifier is trained frame-wise (frame label being identical to subclip label the frame is extracted from). Frames are passed through the backbone and then through a fully connected neural network which output label prediction vector per frame. Once trained, subclip evaluation is performed by a voting system based on each subclip frame classification vectors (average pooling followed by argmax).
- **Configuration 2** (C2): Our original design integrating LSTM module is used. Every single subclip frame passes through the backbone, and the 16 feature vectors are consider as token for the LSTM. Last LSTM module output is then fed into a fully connected network for the classification of the subclip. Training and evaluation follow the same pipeline.
- **Configuration 3** (C3): Video classification models take directly clips as inputs and they are also trained and evaluated on subclips.

C1 stands a baseline for evaluating spatial context value at frame level. Comparison to C2 allows for evaluating the temporal component value in the classification result. C3 stands as more complex architecture embedding spatial and temporal components and comparison to C2 allows for evaluation of the model complexity impact on prediction accuracy.

4. Experiments and results

4.1. Dataset setup

Train, validation and test subsets are respectively constituted of 487, 165 and 134 clips. We have taken particular care to respect the balance of the classes occurrences in between subsets and to prevent a patient from appearing in two different subsets. Indeed, as we are facing diseased parenchyma which may significantly differ from patient to patient in their ultrasound imaging, we took particular attention not to introduce such bias in our analysis and particularly during the generalization testing procedure.

4.2. Experimental setup

All models were pre-trained on ImageNet (Deng et al. 2009), except for models from C3 which were pre-trained on Kinect-400 (Kay et al. 2017), and then finetuned on our dataset. To prevent overfitting, geometric and pixel level data augmentation have been applied. For each model, the checkpoint achieving best results on the validation subset was used for the evaluation on the test subset. The accuracy was calculated for each class separately and averaged across classes with equal weights. We report the top-1 accuracy for all models/configurations in Figure 4. Overall, MViT (Fan et al. 2021) (C3) performs the best on our dataset with top-1 accuracy of 66.8% while XCiT T-LSTM (C2) is ranked second with top-1 accuracy of 66.1%. We note here the significant parameters numbers difference (11.9 millions for XCiT T-LSTM compared to 36 millions for MViT). Note that if we do not distinguish between pancreas parts classes, we reach 80% accuracy for pancreas detection.

We also depict in Figure 5 some confusion matrices results. We clearly see a general good classification for *liver* and $artifacts$. We also note a *body* and *head* good classification for the different models. The most striking point is the non-ability of almost all models to capture the *tail* which is most of the time misclassified as *body*. Only MViT model (Fan et al. 2021) is able to distinguish the *tail* with reasonable performance.

5. Discussion and conclusion

Our results show that models using temporal context (C2 and C3) obtain significantly better results than models based only on static frames (C1). For example, the addition of LSTM (C2) improved the results by an average of 4.7 points and up to 10 points (Figure 4) compared to corresponding C1 configuration. Moreover, we can notice that generally the backbones based on ViT obtain better results than the backbones using CNN. Indeed, in average ViT models obtain an accuracy of 58.6 and 61.6 in C1 and C2 respectively, while the CNN obtain an accuracy of 53.3 and 57.5 in C1 and C2 respectively. Moreover MViT outperformed all CNN based video classifier models (C3). We suspect that the intrinsic behavior of ViT could capture more information on the anatomical context than CNN.

Ill pancreas parenchyma is very complicated to identify for an untrained observer from an EUS video because of the poorly defined organ boundaries, inhomogeneous speckle texture and the overall similarity of the surrounding structures. Many years of practice are necessary for gastroenterologists to efficiently recognize degraded parenchyma in this context. Here, we demonstrate that identifying a short video clip

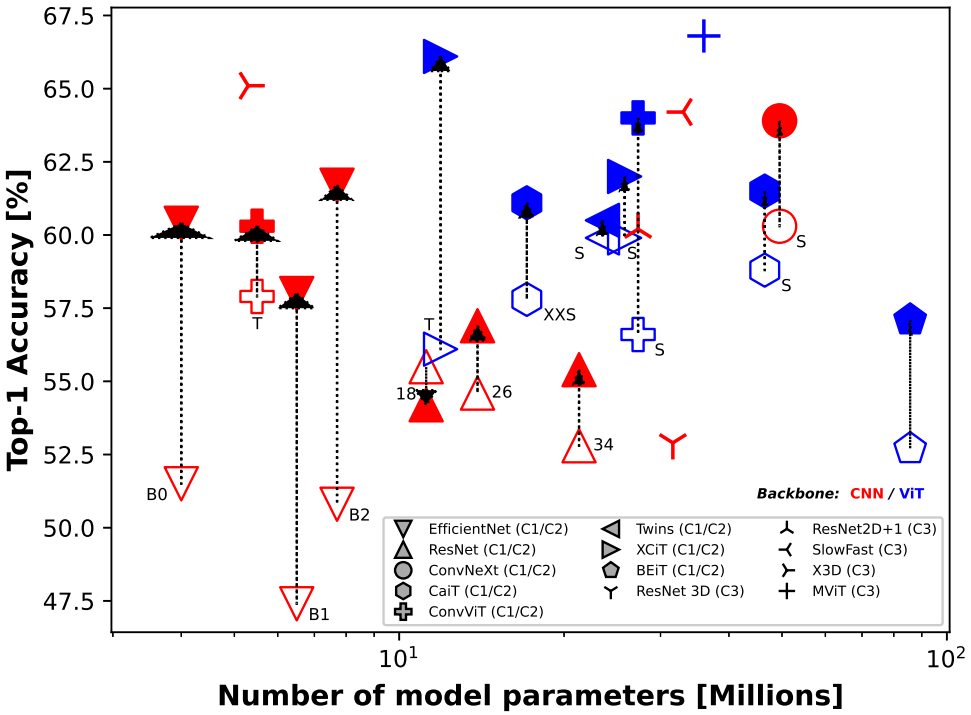

Figure 4. Scatter plot of the accuracy in relation to the number of trainable parameters of the following models and associated configurations: EfficientNet (B0, B1, B2) (Tan and Le 2019), ResNet (18, 26, 34) (He et al. 2016), ConvNeXt (S) (Liu et al. 2022), CaiT (XXS, S) (Touvron et al. 2021), ConvViT (S, T) (d'Ascoli et al. 2021), Twins (S) (Chu et al. 2021), XCiT (S) (Ali et al. 2021), BEiT (Bao et al. 2021), ResNet 3D (Karpathy et al. 2014), ResNet2D+1 (Tran et al. 2018), SlowFast (Feichtenhofer et al. 2019), X3D (Feichtenhofer 2020) and MViT (Fan et al. 2021). CNN and ViT backbone models are resp. red and blue colored. C1, C2 and C3 configurations are resp. marked as empty, plain and wireframe symbols. When needed, backbone *flavor* is indicated near the C1 empty symbol. An arrow is indicating evolution from C1 to C2 configuration accuracies.

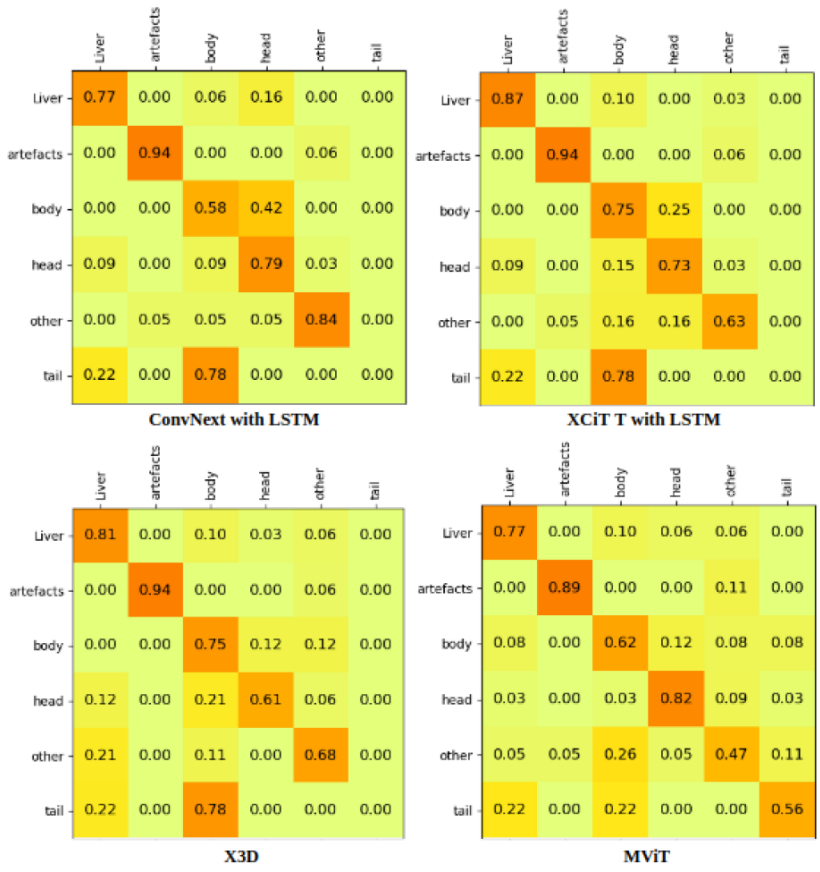

Figure 5. Confusion matrices for different models.

as containing pancreas (degraded) parenchyma is already possible with a confidence higher than 80%. This achievement already holds great potential for clinical application where EUS training practitioners could be guided by an automated system displaying presence/absence of pancreas in the control screen, helping focusing attention when necessary while manipulating the flexible endoscope in the meantime. This would significantly reduce mental workload and accelerate EUS learning curve. We also point out here that such value is typically what intermediate EUS users in our team can achieve during annotation work where video slow down, replay and full video analysis is possible at user pace with "only" annotation task under consideration. However, this is not characteristic of relevant medical scenario when practitioner also manipulates the endoscope, search for parenchyma tissue in US video in order to locate pancreas part under consideration and screens for potential lesions in the same time, all of this at "real" video speed. So we believe this performance in classification is already valuable in clinical use for beginner to intermediate users where real-time analysis with comparable efficiency could guide user during his complex pancreas EUS screening.

When considering pancreas parts, our study shows very promising results with overall classification as high as 66.8%. Looking at confusion matrices to refine our interpretation, it appears that C2 and C3 models generally led to very good classification of classes *head*, *artefact*, *liver* and *other*. However, this is significantly different for the *tail* that is always classified as *body* except for MViT which demonstrates reasonable performances with classification accuracy as high as 56% for this label. As we explained earlier, this misclassification error is not surprising considering the high similarity between body and tail segment. It is also one of the most difficult tasks even for endoscopists. But MViT's significant results in this tail/body discrimination is highly interesting. Our best hypothesis here is that its multi-scale architecture is at the origin of its better performance. This tail classification issue is crucial for our application as it is a highly important marker for ensuring that the gland screening has been done properly and there was no unexamined part at the extremity of the pancreas. So efforts still have to be devoted to understand and improve its classification in EUS videos. Nevertheless, if we consider the relatively low volume represented by the tail, and the subsequent low fraction of EUS video focused on this segment, we can conclude that our approach is already very efficient to classify most of an EUS video.

In this paper, we present a new database of pancreas segments and liver annotated EUS videos. We have evaluated 35 models and demonstrated the superiority of temporal models when combined with the ability of vision transformers to treat ultrasound images. This pilot study shows very encouraging results and great promise in assisting gastroenterologists in EUS with direct impact on patient care. In order to test for guidance efficiency in real clinical scenario, we will start a study based on this result in order to check if a pancreas part detection tool, exhibiting such level of performance in real-time, could significantly help practitioner in their daily task in pancreatic EUS screening.

Funding

This work was carried out within the framework of the project APEUS supported by the ARC Foundation (www.fondation-arc.org) and was partially supported by French state funds managed within the "Plan Investissements d'Avenir" and by the ANR

(reference ANR-10-IAHU-02).

References

Adler DG, Diehl DL. 2015. Missed lesions in endoscopic ultrasound. Endoscopic Ultrasound. 4(3):165.

Ali A, Touvron H, Caron M, Bojanowski P, Douze M, Joulin A, Laptev I, Neverova N, Synnaeve G, Verbeek J, et al. 2021. Xcit: Cross-covariance image transformers. Advances in neural information processing systems. 34.

Bao H, Dong L, Wei F. 2021. Beit: Bert pre-training of image transformers. arXiv preprint arXiv:210608254.

Białek EJ, Jakubowski W. 2017. Mistakes in ultrasound diagnosis of superficial lymph nodes. Journal of ultrasonography. 17(68):59.

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. 2020. End-to-end object detection with transformers. Available from: https://arxiv.org/abs/2005.12872.

Chu X, Tian Z, Wang Y, Zhang B, Ren H, Wei X, Xia H, Shen C. 2021. Twins: Revisiting the design of spatial attention in vision transformers. Advances in Neural Information Processing Systems. 34.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee. p. 248–255.

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. Available from: https://arxiv.org/abs/2010.11929.

d'Ascoli S, Touvron H, Leavitt ML, Morcos AS, Biroli G, Sagun L. 2021. Convit: Improving vision transformers with soft convolutional inductive biases. In: International Conference on Machine Learning. PMLR. p. 2286–2296.

Fan H, Xiong B, Mangalam K, Li Y, Yan Z, Malik J, Feichtenhofer C. 2021. Multiscale vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. p. 6824–6835.

Feichtenhofer C. 2020. X3d: Expanding architectures for efficient video recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. p. 203–213.

Feichtenhofer C, Fan H, Malik J, He K. 2019. Slowfast networks for video recognition. In: Proceedings of the IEEE/CVF international conference on computer vision. p. 6202–6211.

Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin D, Piñeros M, Znaor A, Bray F. 2019. Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods. International journal of cancer. 144(8):1941–1953.

Freedman D, Blau Y, Katzir L, Aides A, Shimshoni I, Veikherman D, Golany T, Gordon A, Corrado G, Matias Y, et al. 2020. Detecting deficient coverage in colonoscopies. IEEE Transactions on Medical Imaging. 39(11):3451–3462.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 770–778.

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural computation. 9(8):1735–1780.

Huang J, Lok V, Ngai CH, Zhang L, Yuan J, Lao XQ, Ng K, Chong C, Zheng ZJ, Wong MC. 2021. Worldwide burden of, risk factors for, and trends in pancreatic cancer. Gastroenterology. 160(3):744–754.

Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. 2014. Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. p. 1725–1732.

Kay W, Carreira J, Simonyan K, Zhang B, Hillier C, Vijayanarasimhan S, Viola F, Green T,

Back T, Natsev P, et al. 2017. The kinetics human action video dataset. arXiv preprint arXiv:170506950.

Krizhevsky A, Sutskever I, Hinton GE. 2017. Imagenet classification with deep convolutional neural networks. Commun ACM. 60(6):84–90. Available from: https://doi.org/10.1145/3065386.

Lin TY, Goyal P, Girshick R, He K, Dollár P. 2017. Focal loss for dense object detection. Available from: https://arxiv.org/abs/1708.02002.

Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. 2022. A convnet for the 2020s. arXiv preprint arXiv:220103545.

Redmon J, Farhadi A. 2018. Yolov3: An incremental improvement. Available from: https://arxiv.org/abs/1804.02767.

Tan M, Le Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR. p. 6105–6114.

Tonozuka R, Itoi T, Nagata N, Kojima H, Sofuni A, Tsuchiya T, Ishii K, Tanaka R, Nagakawa Y, Mukai S. 2021a. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: A pilot study. Journal of Hepato-Biliary-Pancreatic Sciences. 28(1):95–104.

Tonozuka R, Mukai S, Itoi T. 2021b. The role of artificial intelligence in endoscopic ultrasound for pancreatic disorders. Diagnostics. 11(1):18.

Touvron H, Cord M, Sablayrolles A, Synnaeve G, Jégou H. 2021. Going deeper with image transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. p. 32–42.

Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. 2018. A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. p. 6450–6459.

Wu L, Cheng JZ, Li S, Lei B, Wang T, Ni D. 2017. Fuiqa: fetal ultrasound image quality assessment with deep convolutional networks. IEEE transactions on cybernetics. 47(5):1336–1349.

Zhang J, Zhu L, Yao L, Ding X, Chen D, Wu H, Lu Z, Zhou W, Zhang L, An P, et al. 2020. Deep learning–based pancreas segmentation and station recognition system in eus: development and validation of a useful training tool (with video). Gastrointestinal endoscopy. 92(4):874–885.

Zhang L, Sanagapalli S, Stoita A. 2018. Challenges in diagnosis of pancreatic cancer. World journal of gastroenterology. 24(19):2047.