Learning More for Free - A Multi Task Learning Approach for Improved Pathology Classification in Capsule Endoscopy

Anuja Vats[1(✉)], Marius Pedersen[1], Ahmed Mohammed[1,2], and Øistein Hovde[1,3,4]

[1] NTNU, Gjøvik, Norway
{anuja.vats,marius.pedersen,mohammed.kedir}@ntnu.no
[2] SINTEF Digital, Smart Sensor Systems, Oslo, Norway
[3] University of Oslo, Oslo, Norway
[4] Innlandet Hospital Trust, Gjøvik, Norway
oistein.hovde@sykehuset-innlandet.no

Abstract. The progress in Computer Aided Diagnosis (CADx) of Wireless Capsule Endoscopy (WCE) is thwarted by the lack of data. The inadequacy in richly representative healthy and abnormal conditions results in isolated analyses of pathologies, that can not handle realistic multi-pathology scenarios. In this work, we explore how to learn more for free, from limited data through solving a WCE multicentric, multi-pathology classification problem. Learning more implies to learning more than full supervision would allow with the same data. This is done by combining self supervision with full supervision, under multi task learning. Additionally, we draw inspiration from the Human Visual System (HVS) in designing self supervision tasks and investigate if seemingly ineffectual signals within the data itself can be exploited to gain performance, if so, which signals would be better than others. Further, we present our analysis of the high level features as a stepping stone towards more robust multi-pathology CADx in WCE. Code accompanying this work will be made available on github.

Keywords: Capsule endoscopy · Multi task learning · Self supervision

1 Introduction

WCE is a diagnostic procedure for screening the lining (mucosa) of the Gastrointestinal Tract (GIT) for abnormalities. It has succeeded in being the first line of diagnosis for the middle part of the GIT [26]. Apart from being widely

preferred for examining the small bowel, due to its non-invasive nature, it is also a preferred alternative to upper endoscopy and colonoscopy [19, 21].

The procedure involves a pill-sized imaging device shaped as a capsule, to be swallowed by the patient. It traverses through the GIT taking pictures of the mucosa and lumen, transmitting them wirelessly to an external data recorder. The data from each capsule is a video stream that is 8 to 16 h long, depending on the capsule model [2, 13, 16] amounting to frames upwards from 50,000 to be investigated by experts. Moreover, a significant number of these frames contribute from little to no diagnostic information due to either debris and bubbles completely or partially occluding the mucosa or undesirable capsule orientations for long duration during its traversal [22].

Owing to its success in recognition tasks, deep learning has led to major breakthroughs in anomaly detection across medical domains. Soffer et al. [23] present a comprehensive review including deep learning based studies in WCE. Most recent approaches apply a Convolutional Neural Net(CNN) based feature extractor for **C**omputer **A**ided **De**tection (CADe) of pathologies like polyps, ulcers, bleeding etc. [9, 20, 23, 26]. A multicentric CADe system trained potentially from millions of images as in [8] can achieve an even higher detection rate than human-readers for certain anomalies. They perform the first of such a large scale multicentric study by rigorously fine-tuning a pre-trained CNN on more than 100 million images. While their method achieves high sensitivity in filtering out suspected anomalies, further attempt to classify the detected anomaly is not as successful. Considering the large amount of multicentre data used in this study, the inability to do so, reveals the complexity involved in reliably representing features that can classify multiple abnormalities beyond obvious confounders. Infact, this happens to be one of the key challenges in CADx arising from the lack of morphological descriptions of pathological and healthy conditions at disposal (the essential difference lies in CADx focusing on features for the purpose of characterizing an abnormality as opposed to localizing an abnormality within a frame (CADe)). Stidham et al. [24] discuss the consequences of such a challenge, with regards to tissue histology. Some of the complicating factors can be understood by considering the response of the GIT under abnormality:

1. It may exhibit similarity in visual characteristics for different conditions and different severity of these conditions.
2. It may exhibit visible differences in pathology appearances that may get overshadowed by other (medically irrelevant but visually pervasive) morphological similarities pertaining to the endoluminal scene (e.g. mucosal surface structures, occlusions, turbidity, illumination etc.)

Our strategy to combat this challenge is to learn more efficiently discriminative features with limited data for free. We achieve this by taking advantage of additional (seemingly ineffectual) teaching signals within data through two ways - self supervision and Multi Task Learning (MTL). The inspiration behind MTL comes from the observation that humans greatly benefit, in their learning, by performing multiple tasks as opposed to learning the same tasks in isolation [7]. By the same logic, neural networks benefit by sharing information across several related tasks, by means of a common underlying representation. Even when the

primary focus is only on one of the tasks, a good choice of additional tasks can lead to better predictive performance by supplementing domain-specific information not necessarily captured by the main task [4,6,7,15,18,27].

Notwithstanding the challenges of WCE, gastroenterologists are usually able to find pathological areas within the image with relative ease. Most lesions and inflammations are obscure and vary considerably in scale considering the global context within images. In addition, the inherent presence of distortions causes further corruption of diagnostically relevant details. Despite this, gastroenterologists can not only identify and localize abnormality but also classify its type. That is, aside from the most difficult of cases [28] the doctor's ability to identify and classify pathologies remains relatively unhindered by distortions. This applies generally to the HVS. Continuing from this argument, if distortions are unavoidable in WCE and doctors can diagnose in spite of them, can they be exploited to provide a useful inductive bias for classification? To this end, we propose a novel MTL framework as a combination of two types of tasks - supervised pathology classification task (SPT) and self-supervised (SS) distortion level classification task (SSDT). The SSDT task is SS as the labels are freely obtained from distorting the images in a pre-calculated way to enable supervised learning from it. Further, we also present feature analysis for WCE CAD-CAP data [14]. To the best of our knowledge, we are the first ones to perform multicentric pathology classification on WCE using MTL and SS. Besides ours, the same classification has been presented in only one other work [25], however, their objectives and approach differ from ours significantly. The main contributions of this paper are

1. We are the first to apply MTL for WCE classification. We combine teaching signals from two different levels of supervision - full supervision (SPT) and self supervision (SSDT) to gain in performance.
2. We increase the pathology classification accuracy by 7% from a Single Task (ST) based pathology classification.
3. We explore the influence of SSDTs and discuss their role in performance.
4. We discover that the clusters of similarity that automatically emerge in a high level feature space correspond to those manually identified in other works [13, 22]. This understanding of the similarity explains the cause for the bottleneck of CADx typical in WCE.

2   Method

Major distortions in WCE are similar but aggravated from those identified in clinical endoscopy [1], owing to the uncontrolled motion of the capsule inside the GIT. In this work, three distortions have been selected for auxiliary SSDT - brightness (B), contrast (C) and motion blur (MB). One MTL experiment is a combination of SPT with one (or two) SSDT. The SSDT comprises of adding a known level of the selected distortion to be predicted onto a centrally cropped, resized (400×400) and randomly rotated image. Each level of distortion (numerical value between (0,3]) corresponds to an output category for SSDT B and C (for more details refer suppl. Task Configuraions). For SSDT MB, we choose to
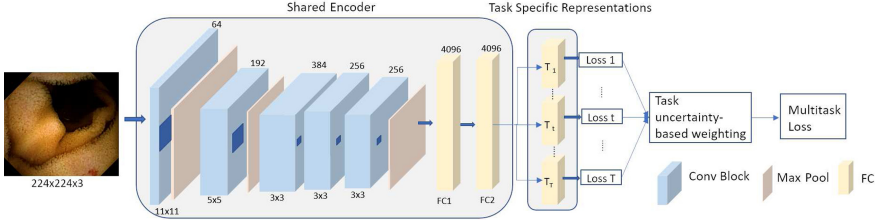
Fig. 1. Overview of our proposed MTL architecture. The encoder holds a shared representation that is simultaneously discriminative for multiple tasks, whereas the task specific layer specializes in a particular task.

add linear motion blur, with output categories corresponding to the direction of added motion [0,45,90,180]. The network at any time does two simultaneous-3 way-pathology and 4 way-distortion level classifications. The training scheme employed is random network initialization with hard parameter sharing [7], under which all layers upto the penultimate task-specific layers share the same representation. This is especially useful when the training data is limited. In having to generalize for multiple tasks using one representation, the network is less prone to overfitting [3]. Figure 1 illustrates the architecture, which has been adapted from AlexNet [12]. Five Convolutional (Conv) and two Fully Connected layers (FC1, FC2) are shared across the tasks whereas, the final FC layers are task specific. A single input image contains signal for both classification tasks at any time.

Objective Function - Every MTL is a joint optimization of two or more losses. In this problem, one of the losses corresponds to SPT while other to SSDT. The total objective conventionally is a weighted sum of individual objectives, in this case a cross-entropy objective for each task. Let $D^t = (x_i, y_i^t), i \in 1, 2, ..., N$ represent the dataset for a task t obtained from a set of tasks given by $t \in 1, 2, ..., T$. (x_i, y_i^t) represent the input output pair for task t such that, in our scheme, each image x_i contains the training signal for all the tasks, and y_i^t is the corresponding task label vector. If $\mathcal{L}_{cet}$ is the cross-entropy loss for task t, then a conventional formulation can be given by

$$\underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{T} \lambda_t \mathcal{L}_{cet}(\theta, f(X^t), Y^t) + \alpha \Omega(\theta) \tag{1}$$

where $f(X^t)$ is the prediction for task t, θ are the network parameters and λ is the task-specific weight. The term $\alpha \Omega(\theta)$ corresponds to regularization concerning network weights with $\alpha \in [0, inf)$. This formulation assumes that optimal weighting of tasks λ_t is either known or can be found. In practice it requires manual tuning which is unintuitive and expensive. We adopt the approach proposed by Kendall et al.[10], where the contribution towards the total loss by a task is learned on the basis of homoscedastic uncertainty. This extends previous formulation for multiple classification tasks as

$$\underset{(\theta,\sigma)}{\text{argmin}} \sum_{t=1}^{T} \left(\frac{1}{\sigma_t^2} \mathcal{L}_{cet}(\theta, f(X^t), Y^t) + log\sigma_t \right) + \alpha\Omega(\theta) \qquad (2)$$

where $\frac{1}{\sigma_t^2}$, the relative contribution of a task to the loss, is an indirect estimate of the uncertainty of task t and learning σ_t can be interpreted as learning the temperature of a Boltzmann distribution [10]. As a task with high uncertainty begins contributing more towards the total loss compared to a task with lower uncertainty, the relative contribution and eventually task uncertainties themselves are optimized as training progresses, that is to say, the optimization in Eq.(2) with respect to both network parameters θ along with σ_t, allows the relative contribution of each task towards total loss to change during the course of training till a reasonably low ratio is reached for all tasks involved. The additional term $log\sigma_t$ is an additional regularizer, penalizing the tendency of σ_t towards larger values.

3   Experiments and Results

The data used for our multicentric pathology classification is from the Computer Assisted Diagnosis for Capsule Endoscopy Database CAD-CAP (GIANA Endoscopic Vision Challenge 2018) comprising of 20,000 normal and 5000 images of varying abnormalities [14]. Out of these, a balanced classification dataset with 1812 images belonging to three classes "normal", "inflammatory lesion" and "vascular lesion" (600, 607, 605 respectively) has been used. 80:10:10 split has been used for train, validation and test sets. In the challenge, the dataset has been used for CADe, while in this paper we use it for CADx. The five MTL configurations studied in this paper are - STL : fully supervised pathology classification (only SPT), MTL (P+X) : refers to (SPT + SSDT of X), where, SSDT of X is the SS auxiliary distortion task of distortion X. X could be Motion Blur (MB), Brightness (B) or Contrast (C). MTL (P+X+Y) is a three task configuration with SPT and a chosen combination of two SSDT of X. We train using Adam optimizer for 300 epochs with a batch size of 64, an initial learning rate of 0.01 and decay rate of 0.1 every 50 epochs on Nvidia Twin Titan RTX.

Results and Analysis - Table 1 shows the results of our experiments. Overall MTL (P+B) achieves the highest accuracy, but on account of maximizing the correct classification of normals, with no improvement (in fact worse discrimination as compared to STL) for the pathology classes. MTL (P+C) is the best with respect to pathology sensitivity, i.e. while the classification of normals is only slightly better than STL, change in image contrast enables pathologies to be better discriminated. MTL (P+MB) is marginally better than STL.

We observed that combining more than one SSDT with SPT (MTL P+B+C) does not further improve the overall performance, (Table 1, suppl. All MTL configurations). This result is not unexpected even when combining fully supervised

Table 1. Performance of STL and different MTL configurations. P_i and P_v refer to the pathology class of inflammatory and vascular lesions respectively. N refers to the class of normals.

Config	Accuracy (%)		Sensitivity (SPT)			Specificity (SPT)		
	Pathology	Distortion	P_i	N	P_v	P_i	N	P_v
STL	51.9	–	0.44	0.72	0.44	0.41	0.68	0.47
P+MB	52.1	40	0.45	0.70	0.38	0.49	0.57	0.44
P+C	57.1	80.95	**0.56**	0.75	**0.48**	0.55	0.68	0.53
P+B	**58.7**	85.7	0.42	**0.88**	0.35	0.46	0.74	0.39
P+B+C	57	73.5, 88.8	0.55	0.79	0.47	0.59	0.55	0.39

tasks [11]. However, compared to STL the gain by combining two tasks comes both in terms of overall accuracy (from correctly classifying normals as in the brightness task) and discriminating between pathologies (higher sensitivity from the contrast task). That is, this task combination can be seen as a way to gain a specific advantage from each of the two auxiliary SSDTs.

Further, we analyse how different MTL configurations impact the learning of the pathology task (SPT). Figure 2(a, b) show the training accuracy of SPT with and without MTL, (c) the plots for $\frac{1}{\sigma_t}$.
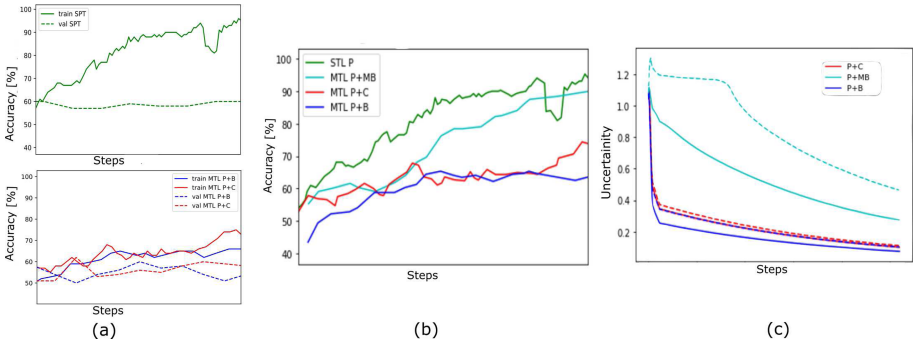

Fig. 2. Impact of SSDTs on the training performance of SPT. (a-top) Overfitting in SPT when trained as a supervised task. (a-bottom, b) Training of SPT under MTL, reduced overfitting with longer training time (c) Square-root of task-uncertainities $\frac{1}{\sigma_p^2}$, $\frac{1}{\sigma_d^2}$ given by $\frac{1}{\sigma_p}(solid)$, $\frac{1}{\sigma_d}(dashed)$ is shown in plot.

It can be seen that, in isolation SPT has a high tendency towards overfitting with quickly escalating training accuracy but plateauing validation accuracy (Fig. 2(a)). The addition of an auxiliary SSDT is seen to reduce this effect (Fig. 2(a, b)). This is a well known advantage of MTL [3] as an overall regularizer. This effect is directly linked to performance seen before, with brightness,

contrast and motion blur in order of decreasing overall accuracy. This is also correlated with the trend in uncertainties. The nature of helpfulness of a task towards pathology can be inferred from how the uncertainties behave during training. From Fig. 2(c), for a helpful SSDT, both uncertainties for SPT $(\frac{1}{\sigma_p})$ and SSDT $(\frac{1}{\sigma_d})$ reduce in tandem, over the course of iterations as more and more confidence is gained from the task. In a relatively unhelpful task (MTL P+MB) however, higher uncertainties indicate this lack of confidence in the task, towards outputs. More specifically, choosing an auxiliary task more complicated (MB) than the main task (SPT) (as seen from the low distortion accuracy for MB in Table 1) can lead from little to no benefit from MTL. This complexity arising in the blur task may be due to the varying amount of inherent linear, rotational and motion blur in the images, that may make prediction of the exact level, especially difficult.

We further investigate the design consideration for an already beneficial task to remain beneficial. Figure 3 shows the training accuracy for two different design choices for SSDT C and B. Config 1 which results in better performance for both tasks is a task of higher complexity (smaller difference between levels of distortion, for this experiment this difference was found to be 0.1 or less for both SSDT B and C) compared to Config 2 (larger difference between levels, 0.3 or more) that results in poorer performance (no gain from STL). This may seem to be in contradiction to the observation before, however it is not. For an MTL to be beneficial the auxiliary task must neither be too easy nor too hard, as in either case, a plateauing effect is observed. MTL P+MB is a difficult task even for simpler designs (larger kernel size being simpler design as the added blur becomes increasingly easier to detect with larger kernels, and vice versa), however, for SSDT C and SSDT B which are tasks of lower complexity with regards to pathology, the plateauing occurs when they are made simple by design. This is done by increasing discrimination between the levels that correspond to output categories (refer Fig. 3, suppl. Task Configurations), In other words, since the auxiliary task is too easy, the distortion task has very high accuracy from the very beginning of the training process. Coupled with a difficult task (SPT), an easy SSDT loses its primary advantage of helping the pathology task out of early local minimas. These findings regarding the nature of beneficial tasks coincide with the observation from [5]. Figure 3(c) shows the task uncertainties during training.

A Note on Similarity - In this section, our focus is to answer the question about the correspondence between *representational similarity* and *medically-perceived similarity*. Medical similarity can be simply described as - two images having the same pathology are similar and different from all images with different pathology. Representational similarity on the other hand is the similarity as understood by the network on the basis of extracted features. A disagreement
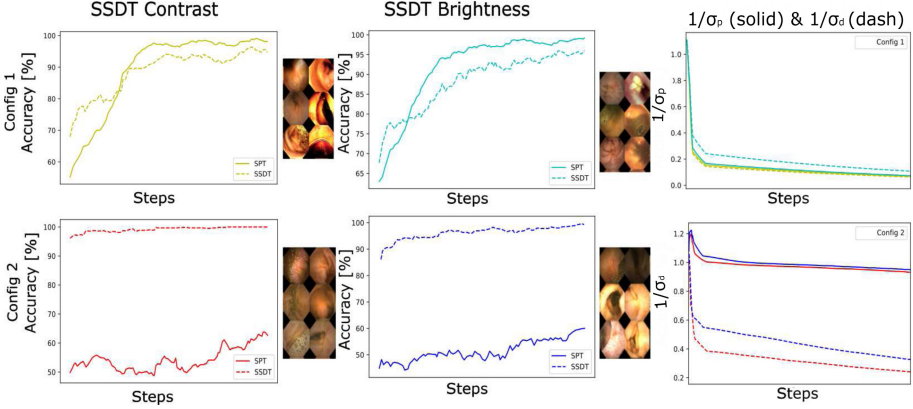
Fig. 3. Comparison of two different SSDT designs. Config 1. improves SPT beyond STL performance and vice versa for Config 2. Pre-processed images, adjacent to the graphs illustrate the distortion levels for the two configs.

between these two reveals the cause for confusion in the network. Although MTL is less prone to this than STL, it still is a major bottleneck in performance. We projected the features from FC2 (Fig. 1) on 3D after dimension reduction using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) [17]. UMAP helps us interpret not only, what features dictate which images are most similar with respect to each other (local), but also what two features are more similar compared to others (global).

As seen in Fig. 4, MTL reveals the inherent variations within the dataset. The extracted features are such that while images with the same pathology lie close (even when in different clusters), pathology does not dictate similarity everywhere. The images within a cluster have representational similarity, that may be medically unimportant. In this work, we saw five such dominant factors emerge out of the feature space repeatedly - bubbles and turbidity, folds in mucosa, lumenal view, clear normal, clear pathology (these were verified by a medical expert to be similarities typical to endoluminal structure or scene, except for few clusters of clear pathology where pathology was the dominant feature). Moreover, at global scale, the feature "bubbles and turbidity" appears more dissimilar from all other features in all experiments. Images in between clusters show mingled characteristics. Similar dominant features have been used by other authors in works before [13,22]. We stress at this point that, unlike other works, these features emerge automatically for us as a result of the learning strategy applied and training under STL does not provide this advantage (refer suppl. UMAP for STL for more details).
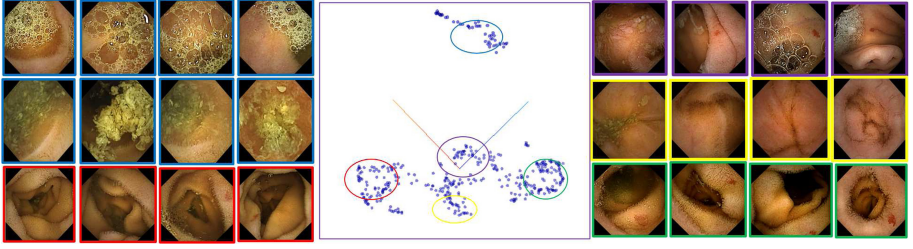
Fig. 4. UMAP of features from FC2 in MTL P+C. Each set of neighbors reveal a dominant similarity. Bubbles and turbid (blue), mucosal folds (red), clear normal (yellow), lumen (green), clear pathology (violet) (Color figure online)

4   Conclusion

In this paper, we demonstrate an MTL approach to gain performance over supervised WCE classification with no added data and a simple CNN architecture, purely by exploiting signals inherent in data under self supervision, additionally to supervision. Further, forcing the network into sharing features between multiple tasks reveals a hierarchy of features in WCE whereby factors that dictate similarity are not inspired only by the presence or absence of pathology, rather are typical of the endoluminal scene. Hence, images with different pathology can be easily perceived to be similar in the presence of such a factor. Our method can be used as a filter for these irrelevant salient patterns, allowing a better subsequent characterization of pathology in CADx. We also experimentally correlate task uncertainties with the nature (beneficial and non beneficial) of tasks [5] which can be used as a way to choose good auxiliary tasks for MTL. This framework is easily scalable to more complex architectures and auxiliary tasks.

References

1. Ali, S., et al.: Endoscopy artifact detection (EAD 2019) challenge dataset. arXiv preprint arXiv:1905.03209 (2019)
2. Atsawarungruangkit, A., Elfanagely, Y., Asombang, A.W., Rupawala, A., Rich, H.G.: Understanding deep learning in capsule endoscopy: can artificial intelligence enhance clinical practice? Artif. Intell. Gastrointest. Endosc. **1**(2), 33–43 (2020)
3. Baxter, J.: A bayesian/information theoretic model of learning to learn via multiple task sampling. Mach. Learn. **28**(1), 7–39 (1997)
4. Benton, A., Mitchell, M., Hovy, D.: Multitask learning for mental health conditions with limited social media data. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, vol. 1, Long Papers, pp. 152–162 (2017)
5. Bingel, J., Søgaard, A.: Identifying beneficial task relations for multi-task learning in deep neural networks. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, vol. 2, Short Papers, pp. 164–169. Association for Computational Linguistics, Valencia, Spain, April 2017. https://www.aclweb.org/anthology/E17-2026

6. Caruana, R.: Multitask learning. Mach. Learn. **28**(1), 41–75 (1997). https://doi.org/10.1023/A:1007379606734

7. Caruana, R.: Multitask learning: a knowledge-based source of inductive bias. In: Proceedings of the Tenth International Conference on Machine Learning, pp. 41–48. Morgan Kaufmann (1993)

8. Ding, Z., et al.: Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. Gastroenterology **157**(4), 1044–1054 (2019)

9. Hwang, Y., Park, J., Lim, Y.J., Chun, H.J.: Application of artificial intelligence in capsule endoscopy: where are we now? Clin. Endosc. **51**(6), 547–551 (2018)

10. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7482–7491 (2018)

11. Kokkinos, I.: Ubernet: training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6129–6138 (2017)

12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017)

13. Laiz, P., Vitria, J., Seguí, S.: Using the triplet loss for domain adaptation in WCE. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 399–405 (2019)

14. Leenhardt, R., Li, C., Le Mouel, J.P., Rahmi, G., Saurin, J.C., Cholet, F., Boureille, A., Amiot, X., Delvaux, M., Duburque, C., et al.: CAD-CAP: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy. Endosc. Int. Open **8**(3), E415 (2020)

15. Liu, X., Gao, J., He, X., Deng, L., Duh, K., Wang, Y.y.: Representation learning using multi-task deep neural networks for semantic classification and information retrieval. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 912–921. Association for Computational Linguistics, Denver, Colorado, May–June 2015. https://doi.org/10.3115/v1/N15-1092, https://www.aclweb.org/anthology/N15-1092

16. McAlindon, M.E., Ching, H.L., Yung, D., Sidhu, R., Koulaouzidis, A.: Capsule endoscopy of the small bowel. Ann. Transl. Med. **4**(19), 369 (2016)

17. McInnes, L., Healy, J., Melville, J.: UMAP: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)

18. Misra, I., Shrivastava, A., Gupta, A., Hebert, M.: Cross-stitch networks for multi-task learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3994–4003 (2016)

19. Mohammed, A., Farup, I., Pedersen, M., Hovde, Ø., Yildirim Yayilgan, S.: Stochastic capsule endoscopy image enhancement. J. Imaging **4**(6), 75 (2018)

20. Muhammad, K., Khan, S., Kumar, N., Del Ser, J., Mirjalili, S.: Vision-based personalized wireless capsule endoscopy for smart healthcare: taxonomy, literature review, opportunities and challenges. Future Gener. Comput. Syst. **113**, 266–280 (2020)

21. Park, J., Cho, Y.K., Kim, J.H.: Current and future use of esophageal capsule endoscopy. Clin. Endosc. **51**(4), 317–322 (2018)

22. Seguí, S., Drozdzal, M., Pascual, G., Radeva, P., Malagelada, C., Azpiroz, F., Vitrià, J.: Generic feature learning for wireless capsule endoscopy analysis. Comput. Biol. Med. **79**, 163–172 (2016)

23. Soffer, S., Klang, E., Shimon, O., Nachmias, N., Eliakim, R., Ben-Horin, S., Kopylov, U., Barash, Y.: Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. Gastrointest. Endosc. **92**(4), 831–839 (2020)
24. Syed, S., Stidham, R.W.: Potential for standardization and automation for pathology and endoscopy in inflammatory bowel disease. Inflamm. Bowel Dis. **26**(10), 1490–1497 (2020)
25. Valério, M.T., Gomes, S., Salgado, M., Oliveira, H.P., Cunha, A.: Lesions multiclass classification in endoscopic capsule frames. Procedia Comput. Sci. **164**, 637–645 (2019)
26. Yang, Y.J.: The future of capsule endoscopy: the role of artificial intelligence and other technical advancements. Clin. Endosc. **53**(4), 387–394 (2020)
27. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multitask learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 94–108. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_7
28. Zheng, Y., Hawkins, L., Wolff, J., Goloubeva, O., Goldberg, E.: Detection of lesions during capsule endoscopy: physician performance is disappointing. Am. J. Gastroenterol. **107**(4), 554–560 (2012)