scientific reports

OPEN

Multi-task deep learning for glaucoma detection from color fundus images

Lucas Pascal[1,2], Oscar J. Perdomo[3], Xavier Bost[2], Benoit Huet[4], Sebastian Otálora[5,6] & Maria A. Zuluaga[1,6]✉

Glaucoma is an eye condition that leads to loss of vision and blindness if not diagnosed in time. Diagnosis requires human experts to estimate in a limited time subtle changes in the shape of the optic disc from retinal fundus images. Deep learning methods have been satisfactory in classifying and segmenting diseases in retinal fundus images, assisting in analyzing the increasing amount of images. Model training requires extensive annotations to achieve successful generalization, which can be highly problematic given the costly expert annotations. This work aims at designing and training a novel multi-task deep learning model that leverages the similarities of related eye-fundus tasks and measurements used in glaucoma diagnosis. The model simultaneously learns different segmentation and classification tasks, thus benefiting from their similarity. The evaluation of the method in a retinal fundus glaucoma challenge dataset, including 1200 retinal fundus images from different cameras and medical centers, obtained a 96.76 ± 0.96 AUC performance compared to an 93.56 ± 1.48 obtained by the same backbone network trained to detect glaucoma. Our approach outperforms other multi-task learning models, and its performance pairs with trained experts using ∼ 3.5 times fewer parameters than training each task separately. The data and the code for reproducing our results are publicly available.

Glaucoma is one of the leading causes of irreversible but preventable blindness in working-age populations[1], which relates to an abnormal fluid balance in the eye that causes an increase in internal ocular pressure. The increased pressure gradually damages the eye optic nerve. If not diagnosed, these induced damages may lead to permanent vision loss. In 2020, it affected approximately 11.2 million people[2,3].

While an early diagnosis is critical to prevent irreversible damages, patients affected by glaucoma usually do not present symptoms in the early stages of the disease. It is thus essential to develop inexpensive detection methods to massively and systematically control patients before the symptoms appear. One way to achieve this is by performing a visual examination of the posterior pole or retinal fundus image. Specialized cameras obtain the color fundus images in a short image acquisition time. The analysis of the fundus images is performed by ophthalmologists, where the most discriminant symptom for detecting glaucoma on fundus images is the presence of a "cupping," which is the retraction of the optic disc (OD) on the optic cup (OC). This cupping causes an increase in the vertical Cup-to-Disc ratio (vCDR), which is the height ratio between the OC and OD. Establishing an accurate diagnosis from these images is particularly difficult and prone to error in the accurate estimation of vCDR.

Deep convolutional networks have shown to be beneficial in medical imaging and in tasks of disease classification in eye fundus[4–7], learning relevant features and patterns directly from images. Over the last years, glaucoma detection using deep learning models reached a remarkable performance at the pair with residents in ophthalmology [3,8–11], thus representing a viable alternative to support current visual assessment. However, automating glaucoma diagnosis suffers from lack of data. Existing annotated datasets contain a few hundred samples, while deep learning models require extensive databases to guarantee a good generalization. Moreover, these models include millions of trainable parameters, requiring significant computational resources for training and deployment[12,13].Therefore, it is essential to develop methods that can make the most from the limited

[1]Data Science Department, EURECOM, 06410 Sophia Antipolis, France. [2]Orkis, 13290 Aix-en-Provence, France. [3]School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia. [4]Median Technologies, 06560 Valbonne, France. [5]Support Center for Advanced Neuroimaging, University Institute of Diagnostic and Interventional Neuroradiology, 3010 Bern, Switzerland. [6]These authors contributed equally: Sebastian Otálora and Maria A. Zuluaga. ✉email: maria.zuluaga@eurecom.fr

Models			Tasks			
			Glaucoma	OD	OC	Fovea
Model	#P × 1e6	time	AUC (↑)	DSC (↑)		Fovea Error (↓)
STL	61.2	0.686	93.56 ± 1.48	**95.45 ± 0.20**	86.96 ± 0.56	**2.33 ± 0.33**
MTL-IO (ours)	**17.2**	0.557	**96.76 ± 0.96**	95.24 ± 0.11	**87.45 ± 0.90**	2.94 ± 0.24
Vanilla MTL	**17.2**	**0.251**	94.78 ± 0.61	94.47 ± 0.25	86.45 ± 0.46	5.56 ± 0.44
GradNorm	**17.2**	0.260	93.47 ± 2.00	94.23 ± 0.45	86.23 ± 0.65	5.56 ± 0.64
PCGrad	**17.2**	0.556	91.51 ± 2.44	94.76 ± 0.12	87.09 ± 0.16	7.51 ± 1.37

Table 1. Results for the test set in the REFUGE dataset over the four tasks using 5-fold cross-validation. Best values are in bold.

resources: computational requirements and the available annotated images, thus operating in a low data size regime while guaranteeing a good generalization.

Multi-task learning (MTL)[14] is a learning paradigm that aims to improve generalization by using the domain information contained in the training signals of related tasks as an inductive bias. In practice, this is done by training a shared model for all tasks. In deep MTL, the shared model consists in the parameters of a deep network[15], hence, the resulting model is smaller than having separate networks for each task. Thanks to these features, MTL is a well-suited approach to automated glaucoma detection, where multiple tasks as OC and OD segmentation, and fovea localization are pre-requisite tasks for computer aided diagnosis (CAD) of retinal diseases[16]. The fovea localization task is related to the OD being located from the center of it by about 2-3 times the diameter of the OD[17]. Despite being related tasks, the use of MTL to simultaneously segment the OD and OC, locate the fovea and detect if the image is glaucomatous has not yet been explored, to the best of our knowledge. Instead, current state-of-the-art works treat each task separately through single task models (STL) or propose MTL approaches that do not exploit the full set of available tasks.

Among STL approaches, Cheng et al.[18] proposed a super-pixel-based segmentation of the OD and OC for glaucoma screening, achieving a performance in terms of area under the curve (AUC) of 0.822. Fu et al.[19] obtained 0.899 with a U-Net-based deep learning method and a transformation of the image to polar coordinates. Among the authors that have explored MTL techniques, Mojab et al. proposed a multi-task model for glaucoma detection composed of two modules for OD and OC segmentations and glaucoma prediction[20], obtaining 90.1 of F-score; the authors did not account for the dissimilarity between the distributions of the segmentation and prediction tasks. Chelaramani reported a novel MTL-based teacher ensemble method for knowledge distillation[21]. The proposed method requires a dataset with a variety of different eye pathologies, which may be difficult to obtain in practice.

This work aims to determine if the relation between tasks associated to glaucoma CAD, i.e. OD and OC segmentation, fovea location and glaucoma detection, can be exploited within an MTL framework to improve model generalization and accuracy for glaucoma detection in a low sample size, low computational resources regime. To this end, a deep MTL model is trained to leverage the similarities of the segmentation of the OD and OC tasks, together with localization of the fovea to detect the presence of glaucoma in retinal fundus images. The proposed MTL approach uses a U-Net encoder-decoder convolutional network as a backbone architecture and adapts it to handle the four tasks using independent optimizers (IO) that can simultaneously learn the segmentation and classification tasks. We denote it MTL-IO. We evaluate our method using the Retinal Fundus Glaucoma Challenge (REFUGE) dataset, including 1200 retinal fundus images (400 for training, 400 for validation, 400 for testing) from different cameras and medical centers, achieving better AUC performance than the same network trained for the single task of detecting glaucoma (92.91 ± 0.69 vs 90.09 ± 2.70). Our approach pairs with trained experts[22,23] and uses approximately 3.5 times fewer parameters than training each task separately.

Results

This section presents the experimental results obtained on the REFUGE challenge dataset, comparing the proposed MTL-IO framework in different setups and against different baselines.

Multi-task learning model with independent optimizers.

We compared our proposed MTL-IO approach to the respective single task model (STL) for each of the tasks and with two state-of-the-art multi-task models: GradNorm[24] and PCGrad[25]. GradNorm[24] adaptively balances the losses by gradient normalization, whereas PCGrad is based on estimating the right signs in the independent task gradients to avoid local minima. To gain understanding of the individual contribution of the IO optimization scheme, we also compare our approach to one using the same pipeline, optimized with a standard optimization scheme[14], which we denote Vanilla MTL. All models were trained five times following a 5-fold cross-validation.

Table 1 shows the classification and segmentation results for each model. Performance is measured in terms of the area under the curve (AUC) for the classification tasks, the Dice score (DSC) for the segmentation tasks, and the L2-distance (Fovea Error) for the localization task. The standard deviation is reported for every performance measure. Model size, in terms of number of parameters (#P), and an iteration time (time), which represents the seconds required for a forward and backward pass in the framework, are also reported.

Model	Tasks			
	Glaucoma	OD	OC	Fovea
	AUC (↑)	DSC (↑)		Fovea Error (↓)
STL	95.63 ± 0.57	**95.68 ± 0.08**	87.45 ± 0.37	**2.19 ± 0.04**
MTL-IO (ours)	**97.03 ± 0.59**	95.13 ± 0.34	**87.55 ± 0.71**	2.34 ± 0.37
Vanilla MTL	96.64 ± 0.83	95.20 ± 0.24	87.53 ± 0.47	3.05 ± 0.37
Res34-Unet	–	93.86 ± 3.70	85.40 ± 7.46	–
GradNorm	95.04 ± 0.95	94.39 ± 0.31	86.38 ± 0.77	5.42 ± 1.30
PCGrad	95.98 ± 0.58	95.32 ± 0.27	87.33 ± 0.60	3.16 ± 0.30

Table 2. Results for the test set in the REFUGE dataset over the four different tasks using 5-fold cross-validation. The results in this table correspond to the models trained with transfer learning. Res34-Unet contains 25.5×10^6 parameters. Given the multiple stages of this method, we roughly estimate an iteration time of 0.375s consisting of the training phases through deep networks and excluding any pre/post-processing stages. Best values are in bold.

MTL-IO outperforms all other methods in glaucoma detection and OC segmentation, while it ranks second in the OD segmentation and the fovea localization task. In terms of model size, all MTL models use approximately 17.2e6 parameters, making them ∼ 3.5 significantly lighter than the STL baseline, which uses 61.2e6 parameters. We estimate the parameters of STL as the sum of parameters of each single-task learner. In terms of computational time, the cumulative iteration time for STL is ∼ 1.2 times slower than MTL-IO. MTL-IO's training iteration time is comparable to PCGrad, but much slower than GradNorm and Vanilla MTL. This difference is explained by the use of the independent optimizers that incur in a computational overhead, which is compensated by the improved performance.

Multi-task learning model with independent optimizers and transfer learning.

Transfer Learning is a widely adopted method to bias a model with prior knowledge on an input domain and lead it to better generalization on new data. In practice, Imagenet[26] pre-trained models have proven to be profitable on a large majority of vision tasks. In medical imaging, although the input domain is different from the Imagenet domain (natural images), the benefits are still noticeable[27], and particularly appreciated to compensate for the usual lack of training data. Its combination with Multi-Task Learning strategies studied here is thus relevant. As Imagenet only involves image classification, there exists no Imagenet pre-trained model for semantic segmentation. However, it is possible to use a pre-trained VGG-16[28] network for the encoding part of the U-Net in the pipeline, while the decoder is initialized from scratch.

Table 2 presents the results of the different models using a pre-trained encoder. As a reference, we have included a state-of-the-art model proposed for optic disc and cup segmentation[29], specifically designed to use transfer learning in its pipeline. We denote it Res34-Unet, as it uses a modified U-Net structure, based on a ResNet-34[30] architecture. To follow their guidelines[29], we used an encoder pre-trained on the Messidor 2 dataset[31,32]. We observe that MTL-IO improves its performance with the AUC reporting 97.03 ± 0.59 in comparison with 96.76 ± 0.96 of the MTL-IO strategy with weights trained from scratch. Interestingly, MTL-IO shows a slight drop in performance for OD segmentation. The drop, however, is not significant and can be considered within the model's variability. The improved performance across tasks is observed for all the other models (Vanilla MTL, GradNorm and PCGrad).

Ablation study: MTL-IO versus single-task learners.

We investigated in further detail the differences between the proposed multi-task approach and the more standard single-task learner strategy. Figure 1 displays the ROC curves for the glaucoma detection task of STL and MTL-IO. It suggests that the multitask classifier benefits from the related tasks to achieve better performance than the single task of glaucoma detection on all operating points (AUC = 0.968 vs. 0.936).

We also analyzed the sensitivity of MTL-IO to different learning rates at training. Figures 2 and 3 respectively show each task best metric score and minimum loss values, for each of the explored learning rates on the validation set. MTL-IO obtains better results over the different learning rates on the glaucoma detection, while the STL model then performs better on the segmentation tasks, and marginally better on the fovea localization task. However, one can notice that it suffers a more important performance drop on the OC segmentation task when evaluating on the test set (see Table 1), suggesting less overfitting for MTL-IO.

Figure 4 shows an example of the segmentation of a Glaucomatous eye. The proposed MTL strategy provides a better segmentation in this challenging case, with a distinctive light dome in the middle of the eye, probably due to poor capture conditions. It is a glaucomatous case, although the vCDR does not suggest it.

Finally, and to foster reproducibility, we assessed the performance of both approaches using the official splits proposed by the REFUGE Challenge , i.e. no cross-validation. Instead, the models were trained three times to account for the variation in weights initialization. Tables 3 and 4 summarize the obtained results with and without the use of transfer learning. We observe a drop in the performance for both STL and MTL-IO, which is explained by the distribution shift observed between the challenge's train and test splits caused by images coming from different imaging devices. In the cross-validation setup, this shift is compensated by the shuffling of

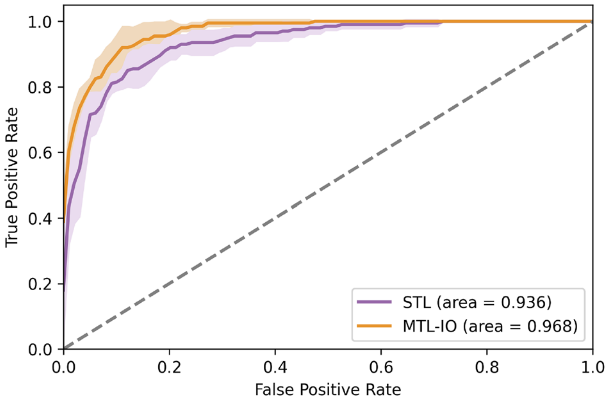

Figure 1. Receiver operating characteristic (ROC) curve for the glaucoma detection task for the single task learning model (STL) and our multi-task learning (MTL-IO) approach.

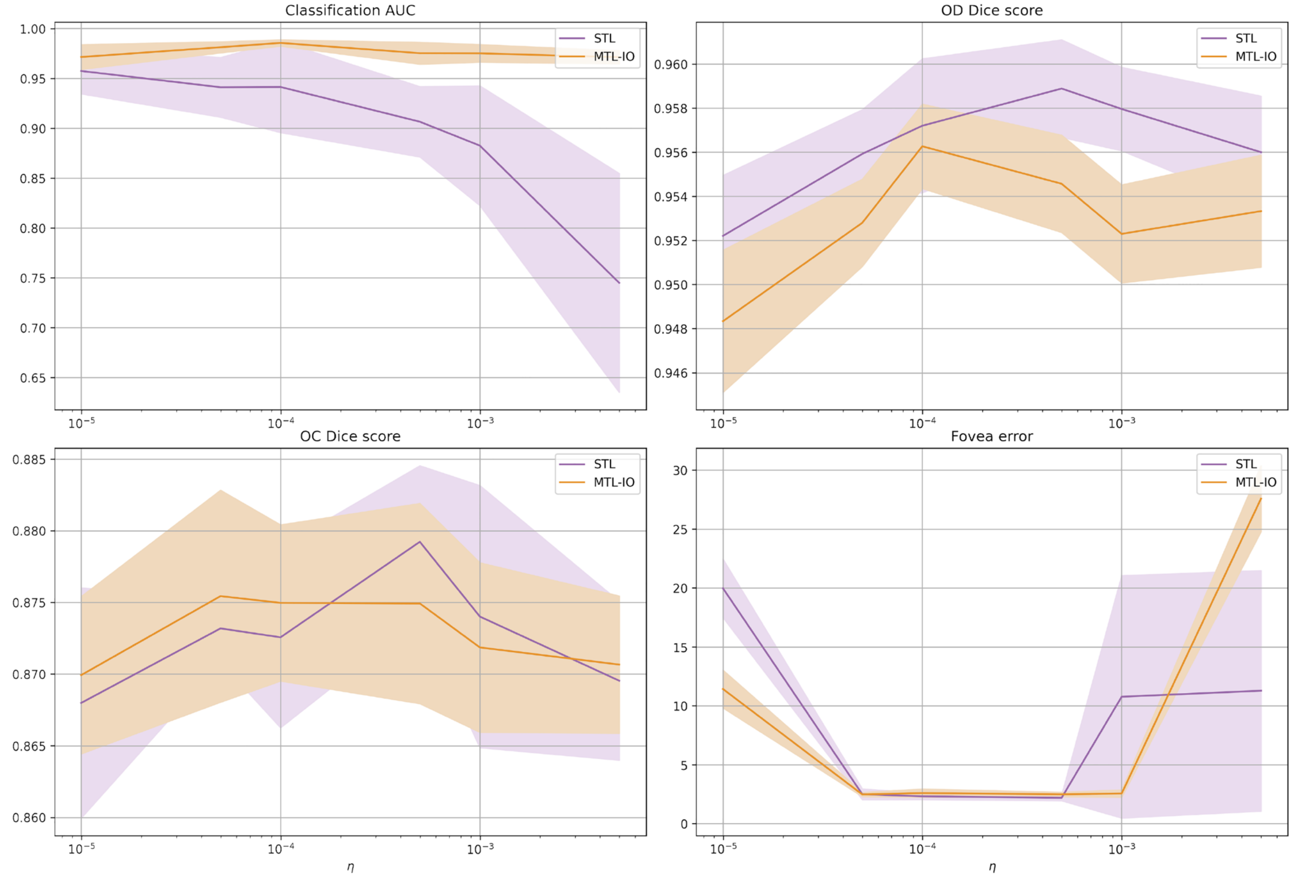

Figure 2. Performance versus learning rate. Per task performance as a function of the learning rate (η) for the single task learning model (STL) and our multi-task learning (MTL-IO) approach. Standard deviation in shaded colour.

the training and validation sets, leading to better results. Despite the drop in performance, MTL-IO remains to be the best performing in terms of AUC.

Discussion

MTL-IO improves generalization by using a unique neural network to learn all tasks jointly. It outperforms all baselines on two of the four proposed tasks , and ranks second behind the STL baseline on the two other tasks while being computationally lighter. Most remarkably, MTL-IO consistently outperforms all baselines on the glaucoma detection task by a large margin.

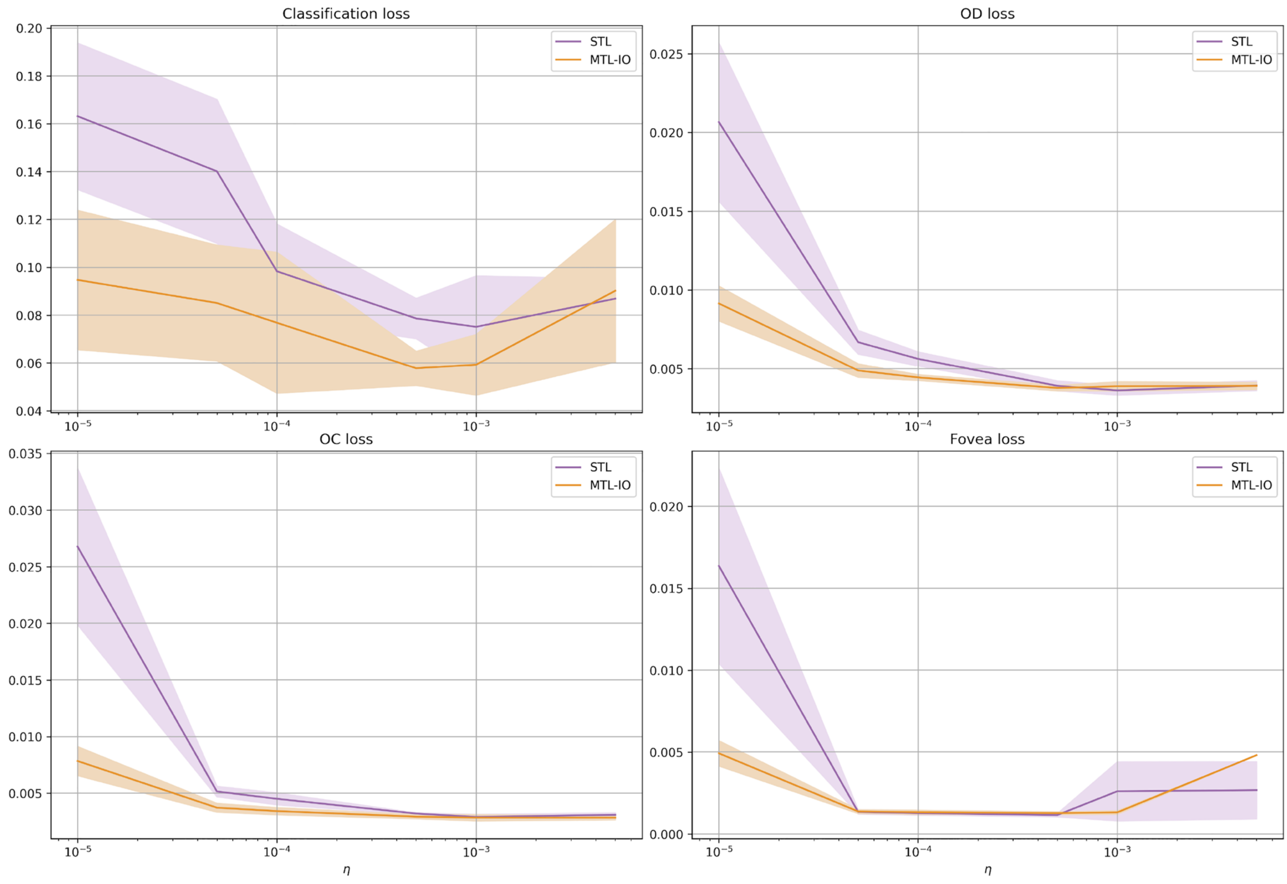

Figure 3. Loss versus learning rate. Final loss values as a function of the learning rate (η) for the single task learning model (STL) and our multi-task learning (MTL-IO) approach. MTL-IO tends to have lower loss values, suggesting the benefit of learning also from other tasks. Standard deviation in shaded colour.

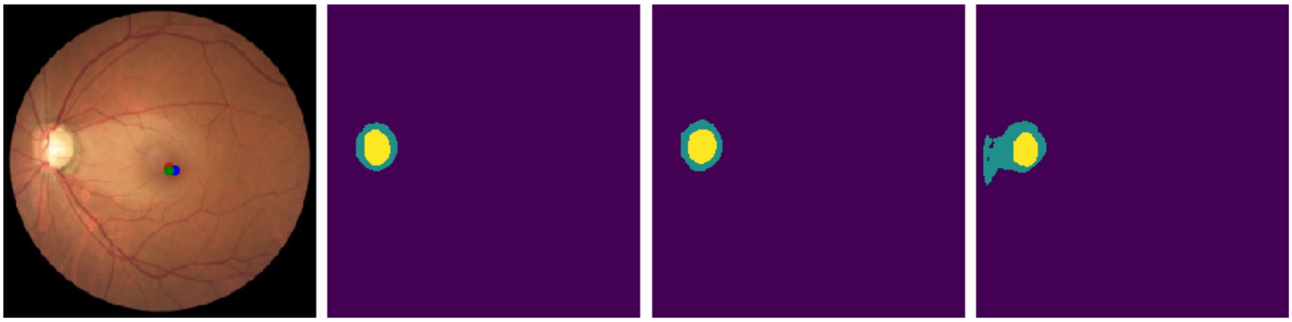

Figure 4. Glaucomatous image from the test set (first image), where the three dots represent the fovea location's ground truth (red), the MTL-IO prediction (blue) and the STL prediction (green), followed by OD (green) and OC (yellow) ground truth (second), MTL-IO (third) and STL (fourth) segmentation masks.

	Tasks			
	Glaucoma	OD	OC	Fovea
Model	AUC (↑)	DSC (↑)		Fovea Error (↓)
STL	87.37 ± 1.51	**93.87 ± 0.74**	80.62 ± 1.25	5.96 ± 0.17
MTL-IO (ours)	**92.61 ± 0.38**	91.61 ± 0.64	**81.21 ± 0.99**	**5.71 ± 0.24**

Table 3. Results for the test set in the REFUGE dataset over the four tasks using the challenge's official splits for training, validation and testing. Best values are in bold.

Model	Tasks			
	Glaucoma	OD	OC	Fovea
	AUC (↑)	DSC (↑)		Fovea Error (↓)
STL	94.30 ± 1.68	**95.29 ± 0.01**	**85.86 ± 0.21**	5.42 ± 0.06
MTL-IO	**96.15 ± 0.14**	94.24 ± 0.38	83.95 ± 0.90	**5.22 ± 0.18**

Table 4. Results for the test set in the REFUGE dataset over the four different tasks with transfer learning using the challenge's official splits for training, validation and testing. Best values are in bold.

When comparing the performance of multi-task and single-task models, it is interesting that the other state of the art MTL methods GradNorm[24] and PCGrad[25] perform worse than the single task baselines on every task, highlighting a task interference issue. Instead, when using the proposed MTL-IO optimization scheme, the multi-task network can significantly reduce task interference and often improve performances compared to the single-task baselines.

In addition to the improved performance, MTL-IO has the advantage that it uses a unique convolutional network for all tasks. This means that it achieves a good performance while being more lightweight than single-task learners: STL is ∼ 3.5 times larger in terms of parameters and ∼ 1.2 times slower than MTL-IO. This is an important feature for real-world use, where resources are often constrained. Our experiments combining transfer learning suggested that the gains achieved by MTL-IO, both in terms of generalization performance and computational efficiency hold in smaller proportions. Although STL observes larger improvements, the MTL-IO remains the best performing at glaucoma detection, which is the main task. As such, it is possible to say that the two strategies, MTL and transfer learning, can be efficiently combined in real-world contexts to create better generalization performance on problems involving multiple tasks.

Despite the above-mentioned advantages, a disadvantage of MTL strategies relates to the extra effort that may be required from a user/expert to put them in place. While an STL strategy requires simple binary labels for training (i.e. presence or absence of glaucoma), MTL techniques also need pixel-wise annotations of the objects to segment and the location of the fovea. All of these annotation tasks are more time consuming and costly. In such setup, it is therefore necessary to assess what is the most critical criterion to optimize. If access to experts for image annotation is difficult, an STL classifier should be used. Instead, if lack of data and limited resources are an important constraint, MTL techniques should be favored.

Materials and methods

Materials: REFUGE challenge dataset.

In 2018 the *Retinal Fundus Glaucoma Challenge* (REFUGE) was launched as a satellite event at the 2018 MICCAI conference. For this event, 1200 retinal fundus images (400 for training, 400 for validation, 400 for testing) from different cameras and medical centers have been collected and annotated by human experts. Annotations were provided for four different tasks: glaucoma diagnosis, optic disc segmentation, optic cup segmentation, and fovea localization. For the diagnosis task, the ground truth is provided as binary labels, attesting to the presence of glaucoma. In the segmentation tasks, the regions defined by the OD (optic nerve head) and the OC (the white elliptic region located inside the optic disc) are provided as binary segmentations. In the fovea localization case, the ground truth is given as the fovea's (x, y) pixel location. All the methods developed and experiments were carried out in accordance with the relevant guidelines and regulations associated to this publicly available dataset.

Methods.

In the following we describe the overall MTL deep learning architecture adopted, the loss functions used for each task and, finally, the independent optimizer (IO) strategy adopted.

Multitask deep learning architecture. We use a U-Net[33], an encoder-decoder convolutional network, with a VGG-16[34] structure and added skip connections between equivalent depths of encoder and decoder, which allow the decoder to recover fine-grained details through the multiple upscalings. This network is well known for solving efficiently biomedical segmentation tasks[35]. Although many variants of the U-Net architecture have been refined for different applications[36], we choose to use its primary version using a VGG16 architecture, as it is the most widely used, and constitutes a default choice for most applications[36–39]. Our MTL approach uses this architecture for two segmentation tasks (OD and OC), one regression task (fovea coordinates) and one classification task (glaucoma diagnosis). The design of the MTL architecture is shown in Fig. 5, and detailed in the following.

Optic disc and cup segmentation tasks. The OD and OC segmentation masks are obtained through the convolutional layer after the shared decoder for each task. Similar to existing works[9], the segmentations of OD and OC are refined through a post-processing step that keeps the main connected component in the prediction map to remove possible prediction noise around these elliptic regions.

Fovea localization. The fovea localization task is addressed as a segmentation task: from the ground truth coordinates of the fovea, a map is created, the center of such map represents the localization of the fovea. The map is a multivariate normal distribution centered in the coordinates (equal variances and null covariances). An example is shown in Fig. 6 (right). The network is trained to fit the maps with a task-respective convolutional

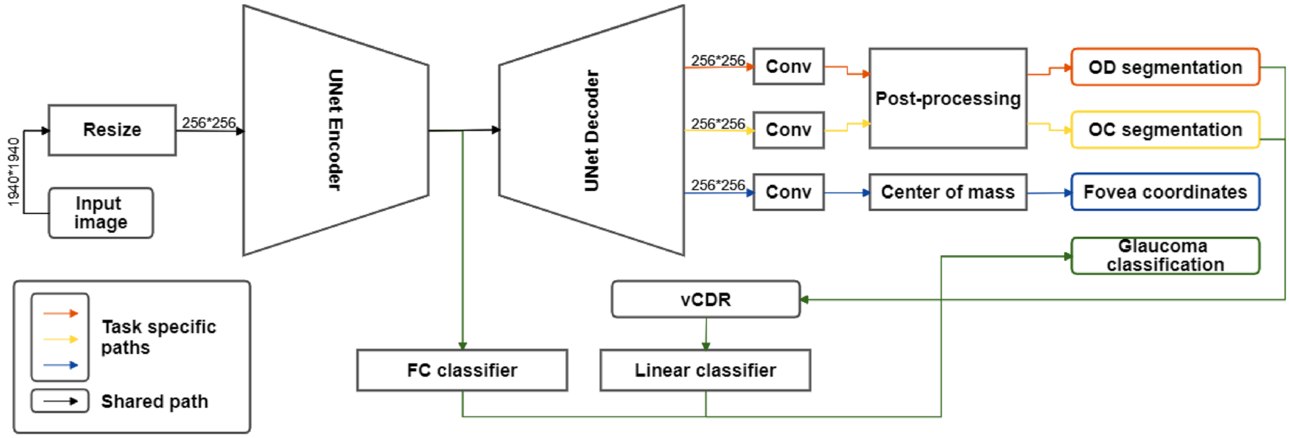

Figure 5. Multi-task learning framework for glaucoma detection, OD and OC segmentation, and fovea localization. The framework uses a U-net as its backbone architecture.

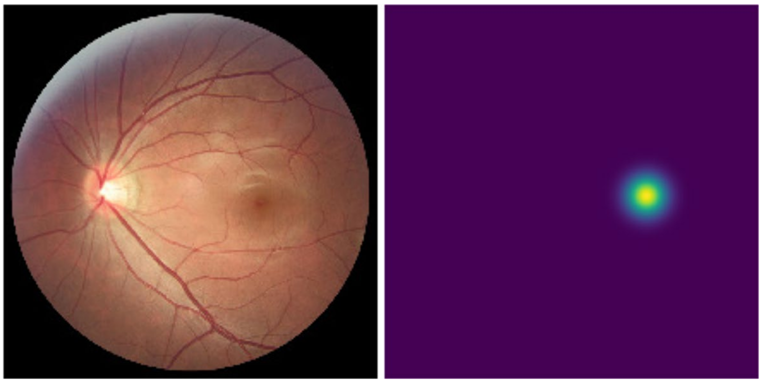

Figure 6. Example of a retinal fundus image (left and the correspondent saliency map centered on the fovea coordinates (right).

layer on the shared decoder. The fovea coordinates are then predicted as the center of mass of the predicted saliency map. In this case, no refinement or postprocessing is performed since it may shift the center of mass.

Glaucoma detection task. The glaucoma detection task (classification) consists of two steps:

1. A prediction is obtained from a fully connected layer, branched after the U-Net encoder (FC classifier).
2. Similarly to some previous works[9], a second prediction is obtained from a logistic regression classifier (Linear classifier), taking as input the vertical Cup-to-Disc Ratio (vCDR) obtained from the OD and OC segmentation tasks. The vCDR is computed as:

$$vCDR = \frac{OC_{height}}{OD_{height}}$$

with OC_{height} and OD_{height} the heights of the OC and OD, obtained from the segmentation branches.

The outputs before the binary outcome of each classifier are averaged. The final classification is obtained by using a threshold of 0.5 over this average.

Loss functions. Here, we present the loss functions used for the optimization of the different tasks.

OD and OC segmentation. The OD and OC segmentation tasks both use a binary cross-entropy loss (BCE), averaged over every pixel *i* of the segmentation maps:

$$\mathscr{L}_{BCE}(p, y) = -\frac{1}{N_{pix}} \sum_{i=1}^{N_{pix}} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

with p, y and N_{pix} respectively the prediction, ground-truth and number of pixels.

Fovea localization. For the fovea localization task, the network is trained to fit the pre-processed saliency maps with a $L1$-loss, since the map values are not binarized:

$$\mathscr{L}_{L1}(p, y) = \sum_i |y_i - p_i|$$

Afterwards, the predicted fovea location is computed as the center of mass of the predicted saliency map.

Glaucoma classification. For the glaucoma classification task, a focal loss[40] is used to better handle the unbalance between positive and negative samples (only 10% of positives):

$$\mathscr{L}_{Focal}(p, y) = (1 - p_t)^{\gamma} log(p_t)$$

with

$$p_t = \begin{cases} p & \text{if} \quad y = 1 \\ (1 - p) & \text{otherwise} \end{cases}$$

Concretely, this loss multiplies the usual binary cross-entropy term with a classification uncertainty term $(1 - p_t)$ to give more importance to uncertain classifications, i.e., those of low populated classes. We set the hyperparameter γ to 2 in our experiments.

MTL independent optimizer optimization strategy. In the following, we present the IO optimization strategy used in this work. It relies on the alternative optimization scheme, alternating independent gradient descent steps on the different task-specific objective functions, as proposed by Pascal et al.[41]. We detail then main steps leading to this optimization scheme, and refer the interested reader to Pascal et al.[41] for more details.

The standard MTL optimization setup with an aggregated loss[14] can be expressed as:

$$\mathscr{L}(w_t, \xi_t) = \sum_{k=1}^{N} c^{(k)} \cdot \mathscr{L}^{(k)}(w_t, \xi_t)$$

where $\mathcal{L}^{(k)}$ is the loss function associated to k^{th} out of N tasks, w_t the shared parameters, and ξ_t the data sample, at iteration t. $c^{(k)}$ are task-specific weighting coefficients, for which we assume uniform weighting, i.e. $c^{(k)} = 1$. If $g^{(k)}$ denotes the derivative of $\mathcal{L}^{(k)}$ with respect to the shared parameters w, the update rule for w at step $t+1$ using stochastic gradient descent is:

$$w_{t+1} = w_t - \eta_t \cdot \sum_{k=1}^{N} g^{(k)}(w_t, \xi_t) \tag{1}$$

where η_t is the learning rate.

Recent works[15,42,43] propose a variation to the update rule in equation 1, in which alternate independent update steps with respect to the different task-specific loss functions are executed, instead of aggregating all the terms at once. This strategy aims to minimize task interference and, hence improve generalization. The alternate update rule can be expressed as:

$$w_{t+1}^{(k)} = \begin{cases} w_t^{(N)} - \eta_t \cdot g^{(k)}(w_t^{(N)}, \xi_t), & k = 1 \\ w_t^{(k-1)} - \eta_t \cdot g^{(k)}(w_t^{(k-1)}, \xi_t), & \forall k > 1 \end{cases} \tag{2}$$

In this work, we adopt the approach from Pascal et al.[41]. It uses a modified alternate update rule (eq. 2) that allows to use individual optimizers (IO) in the form of individual exponential moving averages for each task, to prevent state-of-the-art optimizers (e.g. Adam) from accumulating and mixing previous gradient descent directions of all the different tasks. The modified update rule can be expressed as:

$$w_{t+1}^{(k)} = \begin{cases} w_t^{(N)} - \eta_t \cdot \hat{m}^{(k)}\left(g^{(k)}(w_t^{(N)}, \xi_t)\right), & k = 1 \\ w_t^{(k-1)} - \eta_t \cdot \hat{m}^{(k)}\left(g^{(k)}(w_t^{(k-1)}, \xi_t)\right), & \forall k > 1 \end{cases} \tag{3}$$

where $\hat{m}^{(k)}$ is a task-specific exponential moving average mechanism. Here, the memory term introduced by $m^{(k)}$ only involves previous updates of task k. Such formulation is equivalent to using one independent optimizer per task, and is therefore denoted as MTL-IO. In this paper, we use MTL-IO to denote the complete pipeline.

Implementation details. All methods were implemented in Pytorch 1.2, and ran on NVIDIA Titan XP graphic cards. Kaming uniform initialization[44] was used for all the baselines, except for network parts initialized with transfer learning. For the 5-fold cross-validation, the validation splits were defined on the merged and shuffled train and validation official splits, while the test split was kept unchanged.

Data availibility

The data used to train our models and run experiments is available, upon registration from the REFUGE Challenge (https://refuge.grand-challenge.org/Home2020/). All code to reproduce the results of this article is available in a GitHub repository (https://github.com/robustml-eurecom/glaucoma_mtl). The code can be anonymously downloaded from the following link: https://github.com/robustml-eurecom/glaucoma_mtl/archive/refs/heads/main.zip.

Received: 14 February 2022; Accepted: 7 July 2022
Published online: 20 July 2022

References

1. Weinreb, R. N., Aung, T. & Medeiros, F. A. The pathophysiology and treatment of glaucoma: A review. *JAMA* **311**, 1901–1911 (2014).
2. Tham, Y.-C. *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology* **121**, 2081–2090 (2014).
3. Shibata, N. *et al.* Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci. Rep.* **8**, 1–9 (2018).
4. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).
5. Christopher, M. *et al.* Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci. Rep.* **8**, 1–13 (2018).
6. Shibata, N. *et al.* Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci. Rep.* **8**, 1–9 (2018).
7. Graziani, M. *et al.* Improved interpretability for computer-aided severity assessment of retinopathy of prematurity. In Medical Imaging 2019: Computer-Aided Diagnosis, vol. 10950, 109501R (International Society for Optics and Photonics, 2019).
8. Chen, X. *et al.* Automatic feature learning for glaucoma detection based on deep learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 669–677 (Springer, 2015).
9. Orlando, J. I. *et al.* Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med. Image Anal.* **59**, 101570 (2020).
10. Hemelings, R. *et al.* Deep learning on fundus images detects glaucoma beyond the optic disc. *Sci. Rep.* **11**, 1–12 (2021).
11. Gheisari, S. *et al.* A combined convolutional and recurrent neural network for enhanced glaucoma detection. *Sci. Rep.* **11**, 1–11 (2021).
12. Justus, D., Brennan, J., Bonner, S. & McGough, A. S. Predicting the computational cost of deep learning models. In 2018 IEEE international conference on big data (Big Data), 3873–3882 (IEEE, 2018).
13. Strubell, E., Ganesh, A. & McCallum, A. Energy and policy considerations for modern deep learning research. *In Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 13693–13696 (2020).
14. Caruana, R. Multitask learning. *Mach. Learn.* **28**, 41–75 (1997).
15. Pascal, L., Michiardi, P., Bost, X., Huet, B. & Zuluaga, M. A. Maximum roaming multi-task learning. In 35th AAAI Conference on Artificial Intelligence, vol. 35, 9331–9341 (2021).
16. Xie, R. *et al.* End-to-end fovea localisation in colour fundus images with a hierarchical deep regression network. *IEEE Trans. Med. Imaging* **40**, 116–128 (2021).
17. Welfer, D., Scharcanski, J. & Marinho, D. R. Fovea center detection based on the retina anatomy and mathematical morphology. *Comput. Methods Programs Biomed.* **104**, 397–409 (2011).
18. Cheng, J. *et al.* Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Trans. Med. Imaging* **32**, 1019–1032 (2013).
19. Fu, H. *et al.* Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* **37**, 1597–1605 (2018).
20. Mojab, N., Noroozi, V., Philip, S. Y. & Hallak, J. A. Deep multi-task learning for interpretable glaucoma detection. In 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), 167–174 (IEEE, 2019).
21. Chelaramani, S., Gupta, M., Agarwal, V., Gupta, P. & Habash, R. Multi-task knowledge distillation for eye disease prediction. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 3983–3993 (2021).
22. Azuara-Blanco, A., Burr, J., Thomas, R., Maclennan, G. & McPherson, S. The accuracy of accredited glaucoma optometrists in the diagnosis and treatment recommendation for glaucoma. *Br. J. Ophthalmol.* **91**, 1639–1643 (2007).
23. Li, Z. *et al.* Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* **125**, 1199–1206 (2018).
24. Chen, Z., Badrinarayanan, V., Lee, C.-Y. & Rabinovich, A. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In International Conference on Machine Learning, 794–803 (PMLR, 2018).
25. Yu, T. et al. Gradient surgery for multi-task learning. arXiv preprint arXiv:2001.06782 (2020).
26. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, 248–255 (Ieee, 2009).
27. Tajbakhsh, N. *et al.* Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **63**, 101693 (2020).
28. Jaderberg, M. *et al.* Spatial transformer networks. *Adv. Neural. Inf. Process. Syst.* **28**, 2017–2025 (2015).
29. Yu, S., Xiao, D., Frost, S. & Kanagasingam, Y. Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Comput. Med. Imaging Graph.* **74**, 61–71 (2019).
30. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778 (2016).
31. Decencière, E. et al. Feedback on a publicly distributed image database: The messidor database. Image Analysis & Stereology **33** (2014).
32. Abràmoff, M. D. *et al.* Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol.* **131**, 351–357 (2013).
33. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, 234–241 (Springer, 2015).
34. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
35. Falk, T. *et al.* U-net—Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
36. Siddique, N., Paheding, S., Elkin, C. P. & Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **9**, 82031–82057 (2021).
37. Rampun, A., Jarvis, D., Griffiths, P. & Armitage, P. Automated 2d fetal brain segmentation of mr images using a deep u-net. In Pattern Recognition, 373–386 (2020).
38. Bijay Dev, K. *et al.* Automatic detection and localization of focal cortical dysplasia lesions in MRI using fully convolutional neural network. *Biomed. Signal Process. Control* **52**, 218–225 (2019).

39. Bousselham, A., Bouattane, O., Youssfi, M. & Raihani, A. Improved brain tumor segmentation in mri images based on thermal analysis model using U-net and GPUs. In Advanced Intelligent Systems for Sustainable Development (AI2SD), 80–87 (2020).
40. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision, 2980–2988 (2017).
41. Pascal, L., Michiardi, P., Bost, X., Huet, B. & Zuluaga, M. A. Improved optimization strategies for deep multi-task networks. arXiv preprint arXiv:2109.11678 (2021).
42. Maninis, K.-K., Radosavovic, I. & Kokkinos, I. Attentive Single-Tasking of Multiple Tasks. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1851–1860 (2019).
43. Bragman, F. J., Tanno, R., Ourselin, S., Alexander, D. C. & Cardoso, J. Stochastic filter groups for multi-task cnns: Learning specialist and generalist convolution kernels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 1385–1394 (2019).
44. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, 1026–1034 (2015).

Acknowledgements

Sebastian Otálora thanks the Colombian science ministry Minciencias for partially funding his Ph.D. studies through the call "756 - Doctorados en el exterior". The French government supports Maria A. Zuluaga through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) (ANR-19-P3IA-0002).

Author contributions

L.P., M.A.Z. and S.O. conceived the methodology and experiments, L.P. conducted the experiments, L.P., O.J.P., S.O. and M.A.Z. analyzed the results, L.P., O.J.P and S.O. wrote the original draft, S.O. and M.A.Z. supervised the work. All authors reviewed, edited, and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.A.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.