Concurrent ischemic lesion age estimation and segmentation of CT brain using a Transformer-based network

Adam Marcus, Paul Bentley, and Daniel Rueckert, *Fellow, IEEE*

Abstract—The cornerstone of stroke care is expedient management that varies depending on the time since stroke onset. Consequently, clinical decision making is centered on accurate knowledge of timing and often requires a radiologist to interpret Computed Tomography (CT) of the brain to confirm the occurrence and age of an event. These tasks are particularly challenging due to the subtle expression of acute ischemic lesions and the dynamic nature of their appearance. Automation efforts have not yet applied deep learning to estimate lesion age and treated these two tasks independently, so, have overlooked their inherent complementary relationship. To leverage this, we propose a novel end-to-end multi-task transformer-based network optimized for concurrent segmentation and age estimation of cerebral ischemic lesions. By utilizing gated positional self-attention and CT-specific data augmentation, the proposed method can capture long-range spatial dependencies while maintaining its ability to be trained from scratch under low-data regimes commonly found in medical imaging. Furthermore, to better combine multiple predictions, we incorporate uncertainty by utilizing quantile loss to facilitate estimating a probability density function of lesion age. The effectiveness of our model is then extensively evaluated on a clinical dataset consisting of 776 CT images from two medical centers. Experimental results demonstrate that our method obtains promising performance, with an area under the curve (AUC) of 0.933 for classifying lesion ages $\leq$4.5 hours compared to 0.858 using a conventional approach, and outperforms task-specific state-of-the-art algorithms.

Index Terms—Brain, computer-aided detection and diagnosis, end-to-end learning in medical imaging, machine learning, neural network, quantification and estimation, segmentation, X-ray imaging and computed tomography.

I. INTRODUCTION

A. Marcus is with the Department of Computing and the Division of Brain Sciences, Department of Medicine, Imperial College London, London SW7 2AZ, U.K. (e-mail: adam.marcus11@imperial.ac.uk).
P. Bentley is with the Division of Brain Sciences, Department of Medicine, Imperial College London, London SW7 2AZ, U.K.
D. Rueckert is with the Department of Computing, Imperial College London, London SW7 2AZ, U.K.
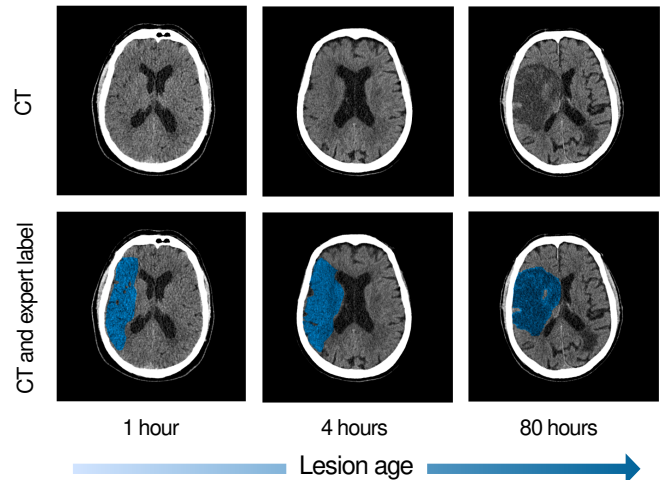

Fig. 1. Example images and expert segmentations of different subjects from our clinical dataset illustrating the appearance of ischemic lesions changing over time.

STROKE is the most frequent cause of adult disability and the second commonest cause of death worldwide [1]. The vast majority of strokes are ischemic and result from the blockage of blood flow in a brain artery, often by a blood clot. Consequently, treatment is focused on rapidly restoring blood flow before irrevocable cell death [2]. The two main approaches are: intravenous thrombolysis, chemically dissolving the blood clot; and endovascular thrombectomy, physically removing the blood clot. Notably, the efficacy of both these treatments decreases over time until their benefit is outweighed by the risk of complications. It is for this reason that current guidelines limit when specific treatments can be given. In the case of thrombolysis, to within 4.5 hours of onset [3]. Therefore, accurate knowledge of timing is central to the management of stroke. However, a significant number of strokes are unwitnessed, with approximately 25% occurring during sleep [4]. In these cases, neuroimaging can help, with previous studies showing promising results using modalities not routinely available to patients, such as magnetic resonance imaging (MRI) and perfusion computed tomography (CT) [5], [6]. Ideally, the widely-available non-contrast CT (NCCT) would be used, but this task is challenging even for detection alone, as early ischemic changes are often not visible to the naked eye (Fig. 1).

A. Related work

Several studies have attempted to delineate early ischemic changes on NCCT. The majority have used image processing techniques or machine learning methods based on hand-engineered features but more recent efforts have used deep learning [7]. Qui *et al.* [8] proposed a random forest voxel-wise classifier using features derived from a pre-trained U-Net and achieved a Dice Similarity Coefficient (DSC) [9] of 34.7% [10]. Barros *et al.* [11] used a convolutional neural network (CNN) and attained a DSC of 37%. El-Hariri *et al.* [7] implemented a modified nnU-Net and reported DSCs of 37.7% and 34.6% compared to two experts. To the best of the authors' knowledge, the current state-of-the-art for this task is EIS-Net [10], a 3D triplet CNN that achieved a DSC of 44.8%.

In contrast, few studies have explored using NCCT to estimate the lesion age. Broocks *et al.* [12] used quantitative net water uptake, originally introduced by Minnerup *et al.* [13], to identify patients within the 4.5 hour thrombolysis time window and attained an area under the receiver operator characteristic curve (AUC) of 0.91. Mair *et al.* [14] introduced the CT-Clock Tool, a linear model using the attenuation ratio between ischemic and normal brain, and achieved an AUC of classifying scans ≤ 4.5 hours of 0.955 with median absolute errors of 0.4, 1.8, 17.2 and 32.2 hours for scans acquired ≤ 3, 3–9, 9–30 and >30 hours from stroke onset. These studies all currently require manual selection of the relevant brain regions, and as of yet, have not utilized deep-learning methods that may allow for improved performance.

Deep learning methods have shown great potential across many domains, with convolutional architectures proving highly successful in medical imaging. Here the inductive biases of CNNs, known to increase sample efficiency [15], are particularly useful due to the scarcity of medical data. However, this may be at the expense of performance, as Transformers [16] have surpassed CNNs across many computer vision tasks. By relying on flexible self-attention mechanisms, Transformer-based models can learn global semantic information beneficial to dense prediction tasks like segmentation but typically require vast amounts of training data to do so. Recently, d'Ascoli *et al.* [15] have attempted to address this by introducing gated positional self-attention (GPSA), a type of self-attention with a learnable gating parameter that controls the attention paid to position versus content and can combine the benefits of both architectures.

Traditionally, Transformer and other deep-learning methods have focused on learning different tasks in isolation with separate networks, yet many real-world problems are naturally multi-modal [17]. This has contributed to the increasing popularity of multi-task learning (MTL) [18], a learning paradigm with the aim of jointly learning related tasks to help improve the generalization performance of all tasks [19]. The underlying insight is that by combining the data of different learning tasks, MTL models can learn robust and universal representations that enable them to be more powerful and reduce the risk of overfitting [19].

B. Motivation

As the appearance of an ischemic lesion is highly dynamic, a perfect segmentation model would need to recognize lesions for all time points and implicitly understand the lesion age. These two tasks, lesion segmentation and age estimation, appear to be inherently complementary, and therefore, it seems reasonable that they may benefit from being learned together.

Furthermore, existing work has shown that estimating lesion age can gain from comparing the affected brain to the spatially distant unaffected side [13]. It may then be particularly advantageous to have a wide receptive field and, consequently, utilizing mechanisms such as a Transformer would seem appropriately suited. However, standard Transformers are typically only effective at large scales, as they lack some of the inductive biases of convolutional neural networks, such as translational equivariance, and thus require large datasets, which are often not available in the medical domain [20]. Hence, it is justifiable to consider architectural modifications and alternative designs of Transformers that have been suggested to address this issue, such as using GPSA modules.

C. Contributions

In this work, we propose a multi-task network to simultaneously perform the segmentation of ischemic lesions and estimate their age in CT brain imaging. The main contributions are: (1) We introduce a novel end-to-end transformer-based network to solve both lesion age estimation and segmentation. To our knowledge, this is the first time a deep learning-based method has been applied to solve the challenging task of estimating lesion age. (2) We enhance the data efficiency of our approach by integrating GPSA modules into the model and using a CT-specific data augmentation strategy. (3) To further improve the performance of our model at estimating lesion age, we suggest a new method to better combine multiple predictions by incorporating uncertainty through the estimation of probability density functions. The effectiveness of the proposed method is then demonstrated by extensive experiments using CT imaging data from 776 patients across two clinical centers.

II. METHOD

A. Network

An overview of the proposed model is presented in Fig. 2. The proposed model is based on the DEtection TRansformer (DETR) panoptic segmentation architecture [21] with modifications to improve sample efficiency, performance, and facilitate lesion age estimation. Table I summarizes the main architectural differences. All activation functions were changed to the Gaussian error linear unit [22] (GELU) and batch normalization [23] was replaced with group normalization [24] to accommodate smaller batch sizes. The main components of the proposed model are: 1) a CNN backbone; 2) a transformer encoder-decoder; 3) lesion age estimation, and bounding box prediction heads; and 4) a segmentation head.

The CNN backbone encoder extracts image features of a 2D CT slice input image. It is comprised of four ResNeXt [26]
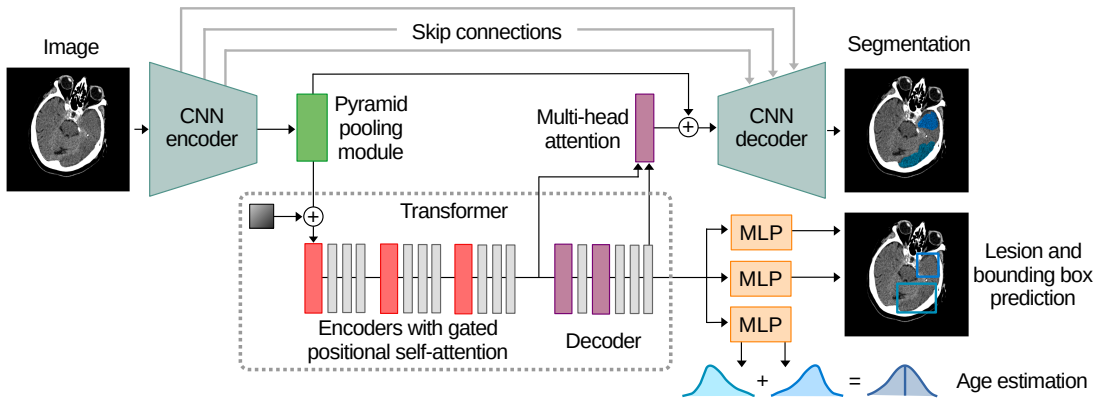
Fig. 2. Overview of the proposed model architecture. Input 3D CT images are processed slice by slice. First, a CNN backbone combined with a pyramid pooling module (PPM) extracts image features at multiple scales. Second, a transformer encoder-decoder with gated positional self-attention (GPSA) uses these features to predict output embeddings for several object queries. Third, multi-layer perceptrons (MLP) use these embeddings to predict lesions, bounding boxes, and lesion age probability distributions. Fourth, a segmentation head generates masks for each lesion based on attention. Finally, the per-slice outputs are combined if the predicted masks are connected in 3D, and for each lesion, the most likely age estimate is used.

TABLE I
SUMMARY OF MAIN ARCHITECTURAL DIFFERENCES BETWEEN DETR
[21] AND THE PROPOSED MODEL.

	Ours	DETR
Activation functions	GELU [22]	ReLU [25]
Normalization layers	GroupNorm [24]	BatchNorm [23]
CNN Backbone	ResNeXt-50 32×4d [26]	ResNet-50 [27]
Feature projection	PPM [28]	1×1 convolution
Attention layers	GPSA [15]	Multi-head attention [16]
Encoder layers	3	6
Decoder layers	1	6
Object queries	10	100
Prediction heads	BBox, class, and regression	BBox and class

GELU = Gaussian error linear unit; ReLU = Rectified linear unit; CNN = Convolutional neural network; PPM = Pyramid pooling module; GPSA = Gated positional self-attention; BBox = Bounding box

blocks and produces an activation map. This activation map is then projected to a feature patch embedding and concatenated with fixed positional encodings [29]. Rather than use a 1×1 convolution as in the original DETR architecture, we use a pyramid pooling module [28] (PPM) that has empirically been shown to increase the effective receptive field by incorporating features extracted at multiple scales.

The transformer encoder-decoder learns the attention between image features and predicts output embeddings for each of the $N = 10$ object queries. Here N was determined by the maximum number of lesions visible in a given slice. We use three transformer encoder blocks and one transformer decoder block following the standard architecture [16] with a couple of exceptions. First, rather than using an auto-regressive model [30] we decode the N objects in parallel. Second, to improve the data efficiency of the model we replace the multi-head attention layers in the encoder with GPSA layers.

The lesion, age estimation, and bounding box prediction heads are each multi-layer perceptrons (MLP) and map the output embeddings of the transformer encoder-decoder to lesion, lesion age, and bounding box predictions. These heads process the queries in parallel and share parameters over all

queries.

The segmentation head generates binary masks for each object instance based on attention. A two-dimensional multi-head attention layer produces attention heatmaps from the attention between the outputs of the transformer encoder and decoder. These heatmaps are then upscaled by a U-Net [31] type architecture with long skip connections between the CNN encoder and decoder blocks.

B. Data Augmentation

To improve the generalizability of our model and prevent overfitting due to limited training data, we adopted a CT-specific augmentation strategy with geometric and appearance transforms. Geometric transforms included: random axial plane flips; $\pm 5\%$ isotropic scaling; $\pm 20\,\mathrm{mm}$ translation; and $\pm 0.5\,\mathrm{rad}$ axial otherwise $\pm 0.1\,\mathrm{rad}$ plane rotation. Appearance transforms included an intensity transform introduced by Zhou *et al.* [32] and a transform we propose to account for the slice thickness variation often present in CT datasets. Regions of the brain are area interpolated [33] to a random slice thickness, ranging from 1–$3\,\mathrm{mm}$ to match the sizes in our dataset, then upscaled back to their original shape. Examples of these transforms can be seen in Fig. 3.

C. Loss Function

We use a combined loss function to enable direct set prediction. The set prediction $\hat{y} = \{\hat{y}_i = \{\hat{p}_i, \hat{b}_i, \hat{s}_i, \hat{a}_i\}\}_{i=1}^{N}$ consists of the lesion probability $\hat{p}_i \in \mathbb{R}^2$ (lesion or no lesion), bounding box $\hat{b}_i \in \mathbb{R}^4$, segmentation mask $\hat{s}_i \in \mathbb{R}^{H \times W}$ where $H \times W$ is the spatial resolution, and lesion age quantiles $\hat{a}_i \in \mathbb{R}^3$ for each of the N object queries. To ensure the loss function is invariant to permutation of the predictions, the Hungarian algorithm [34] was used to assign each instance set label $y_{\sigma(i)}$ to the corresponding query set prediction $\hat{y}_i$ where σ_i represents the best matching order of labels. The combined loss $\mathcal{L}$ is normalized by the number of lesions in a batch and comprises of a lesion loss $\mathcal{L}_p$, bounding box losses $\mathcal{L}_b$ and $\mathcal{L}_g$, segmentation losses $\mathcal{L}_f$ and $\mathcal{L}_d$, and a lesion age loss $\mathcal{L}_a$.
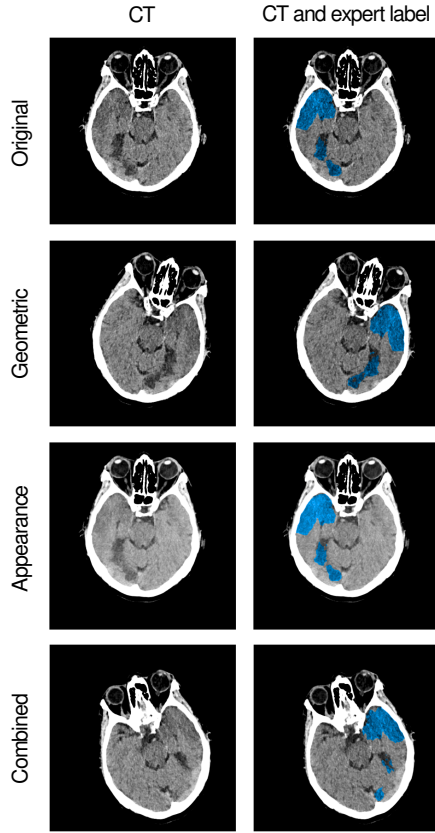
Fig. 3. Example CT-specific geometric and appearance transforms.

$$\mathcal{L} = \sum_{i=1}^{N} (\lambda_p \mathcal{L}_p + \mathbb{1}_{\{p_i \neq \emptyset\}} (\lambda_b \mathcal{L}_b + \lambda_g \mathcal{L}_g + \\ \lambda_a \mathcal{L}_a + \lambda_f \mathcal{L}_f + \lambda_d \mathcal{L}_d)) \quad (1)$$

We used cross-entropy for the lesion loss $\mathcal{L}_p$. For the bounding box losses, L1 loss $\mathcal{L}_b$ and the generalized intersection over union [35] $\mathcal{L}_g$ were used. The segmentation losses comprised of Focal loss $\mathcal{L}_f$ with $\alpha = 0.25$ and $\gamma = 2$ as recommended by Lin *et al.* [36], and Dice loss [37] $\mathcal{L}_d$. To enable the uncertainty of lesion age estimates to be quantized, we used quantile loss for the lesion age loss $\mathcal{L}_a$. We predict three quantiles, assuming that estimates for lesion age are normally distributed, that would correspond to minus one standard deviation from the mean, the mean, and plus one standard deviation from the mean. These can be calculated using ϕ, the cumulative distribution function (CDF) of the standard normal distribution: $\tau_1 = \phi(-1) \approx 0.159$; $\tau_2 = 0.5$; $\tau_3 = \phi(1) \approx 0.841$.

$$\mathcal{L}_a(a_{\sigma(i)}, \hat{a}_i) = \sum_{j=1}^{3} \max(\tau_j ||a_{\sigma(i)} - \hat{a}_{i,j}||_1, \\ (1 - \tau_j)||a_{\sigma(i)} - \hat{a}_{i,j}||_1) \quad (2)$$

In order to account for the varying difficulties of each task common to MTL procedures, we employ a random-weighted loss function where weights are drawn from the Dirichlet distribution [38].

$$\lambda_p, \lambda_b, \lambda_g, \lambda_a, \lambda_f, \lambda_d \overset{\text{i.i.d.}}{\sim} \text{Dir}(1,1,1,1,1,1) \quad (3)$$

D. Inference

At inference time, we combine lesion age estimates if their associated predicted segmentation masks are connected in 3D. Given a set of K lesion age quantile predictions $\hat{a} = \{\hat{a}_k\}_{k=1}^{K}, \hat{a}_k \in \mathbb{R}^3$, we estimate probability density functions (PDF) using $f(x; \mu, \sigma_1, \sigma_2)$, the split normal distribution PDF, where $\mu_k = \hat{a}_{k,2}$, $\sigma_{k,1} = \hat{a}_{k,2} - \hat{a}_{k,1}$, and $\sigma_{k,2} = \hat{a}_{k,3} - \hat{a}_{k,2}$ for each instance.

$$f(x; \mu, \sigma_1, \sigma_2) = \begin{cases} A \exp\left(-\frac{(x-\mu)^2}{2\sigma_1^2}\right) & x < \mu \\ A \exp\left(-\frac{(x-\mu)^2}{2\sigma_2^2}\right) & x \geq \mu \end{cases},$$

where $A = \sqrt{2/\pi}(\sigma_1 + \sigma_2)^{-1}$ \quad (4)

The maximum argument of the sum of these probability density functions is then the combined lesion age estimate, $\hat{a}_\mu$. In the rare instances where a set of predictions produces a negative $\sigma_{k,1}$ or $\sigma_{k,2}$, we resort to the mean lesion age estimate, $\bar{\mu}_k$.

$$\hat{a}_\mu = \underset{x}{\operatorname{argmax}} \sum_{k=1}^{K} f(x; \mu_k; \sigma_{k,1}; \sigma_{k,2}) \quad (5)$$

III. EXPERIMENTS

A. Materials

1) Dataset: Experiments were conducted on a dataset of 776 acute stroke patients with a known time of onset collected across two clinical sites from 2013 to 2019. Extraction and anonymisation of the images followed the pipeline recommended by Muschelli [39]. The median image size was 512 × 512 × 187 voxels with a spatial resolution of 0.45mm × 0.45mm × 0.8mm. Ground truth segmentation masks of 79,959 slices were produced by manual annotation from experts. Lesion ages were calculated using the time from symptom onset to imaging and log-transformed to account for skewed distribution. Patients were randomly divided such that a fixed 20% split were used for testing and the remainder for training and validation using a five-fold group cross-validation approach. Table II lists the characteristics of these groups. When optimizing hyperparameters, 20% of the total dataset was used for validation. An additional independent dataset of 150 patients was collected from the same clinical sites using a similar methodology in order to validate lesion age estimation performance. Instead of producing segmentation masks, experts selected a total of 4,951 lesion containing slices. Full ethical approval was granted by Wales REC 3 reference number 16/WA/0361.

TABLE II
POPULATION CHARACTERISTICS OF THE CLINICAL DATASET

Characteristic	Train and validation set (n = 627)	Test set (n = 149)
Age (years), median (IQR)	74.9 (63.9-82.8)	74.7 (63.1-83.0)
Sex, n (%)		
Male	317 (50.6%)	71 (47.7%)
Female	302 (48.2%)	74 (49.7%)
Missing	8 (1.3%)	4 (2.7%)
ASPECTS, median (IQR)	9 (8-10)	9 (8-10)
NIHSS on admission, median (IQR)	13 (7-20)	13 (7-19)
Affected side, n (%)		
Left	335 (53.4%)	88 (59.1%)
Right	292 (46.6%)	61 (40.9%)
Time from symptom onset to CT (minutes), median (IQR)	232 (109-1212)	253 (110-1325)

IQR = Interquartile range; ASPECTS = Alberta stroke programme early CT score [40]; NIHSS = National Institutes of Health Stroke Scale [41]

2) Evaluation: To evaluate lesion segmentation, we compared mean DSC and intersection over union (IOU) between model predictions and expert segmentation's on a per-subject level. The Mann-Whitney U test was used to determine significance. Identifying the presence of a lesion is more clinically useful than perfect segmentation therefore, we also computed the lesion detection accuracy (LD-ACC), the percentage of expert segmentations $s_{i,j}$ that overlapped with predictions $\hat{s}_{i,j}$ where i and j represent indices for the lesion and image slice, respectively.

$$\text{LD-ACC} = \frac{1}{I} \sum_{i=1}^{I} neq(\sum_{j=1}^{J} s_{i,j}\hat{s}_{i,j})$$

$$\text{where } neq(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases} \quad (6)$$

For lesion age, we excluded subjects with lesions of different ages and calculated the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE). We also evaluated the classification of lesion age within 4.5 hours of onset using accuracy (ACC) and AUC. These were then computed for the regression models by arranging the I total lesions such that $i = 1, ..., I_1$ lesions had an age less than 4.5 hours and $i = I_1 + 1, ..., I$ lesions had an age greater than 4.5 hours, where a_i represents the real and $\hat{a}_i$ the predicted age.

$$\text{ACC} = \frac{1}{I_1} \sum_{i=1}^{I_1} acute(\hat{a}_i) + \frac{1}{I - I_1} \sum_{j=I_1}^{I} 1 - acute(\hat{a}_i)$$

$$\text{where } acute(x) = \begin{cases} 0 & x > \\ 1 & x \leq \log 270 \end{cases} \quad (7)$$

$$\text{AUC} = \frac{\sum_{i=1}^{I_1} \sum_{j=I_1}^{I} S(\hat{a}_i, \hat{a}_j)}{I_1(I - I_1)}$$

$$\text{where } S(x,y) = \begin{cases} 1 & x > y \\ 0.5 & x = y \\ 0 & x < y \end{cases} \quad (8)$$

3) Implementation: All models were implemented using PyTorch [42] version 1.10 and trained from scratch for 100 epochs on a computer with 3.80GHz Intel® Core™ i7-10700K CPU and an NVIDIA GeForce RTX 3080 10GB GPU. The AdamW [43] optimizer was used with a weight decay of 10^{-4}. Learning rate was adjusted from 10^{-6} to 10^{-2} per-epoch using a cyclical schedule [44] and exponentially decayed per-cycle with $\gamma = 0.92$. Gradient clipping [45] was applied to ensure a maximal gradient norm of 0.1. We also employed the stochastic weight averaging [46] for the last 5 cycles. During training, lesion containing regions were linearly sampled from the original volumes to a uniform size, $512 \times 512 \times 1$ for 2D and $128 \times 128 \times 48$ for 3D models, with a spatial resolution of 0.45mm × 0.45mm × 0.8mm. The same CT-specific augmentation strategy was applied for all models. Pixel intensities were clipped based on the 0.5 and 99.5th percentile then normalized using Z-score. Inference of the proposed model required about 14 seconds per subject.

B. Results

1) Comparison with Baseline: We first compare our proposed model to task-specific deep-learning algorithms due to the absence of established methods to jointly perform segmentation and regression. The quantitative results are shown in Table III. For segmentation, we compare against 2D [31], 3D U-Net [47], and TransUNet [48] using the same Focal and Dice loss function. In this task, our proposed model performs slightly better, with significant (p value ≤ 0.05) increases in DSC and IOU at the expense of generally greater computational demands. While the metrics may seem low, it is unlikely to preclude clinical utility due to the high lesion detection accuracies (LDD-ACC), with only 2% of lesions going undetected. Notably, despite the proposed model being 2D in nature, it performed competitively against 3D U-Net, suggesting that for lesion segmentation, the ability to capture global semantic information may outweigh the benefits of learning volumetric relations. These findings are also supported by qualitative evaluation as seen in Fig. 4.

For lesion age estimation, we first trained a linear model based on intensity using a similar methodology to Mair *et al.* [14]. We also trained ResNet-50 [27], ResNeXt-50-32x4d [26], and ConvNeXt-T [49] models using the same quantile loss function. Compared to these models, our proposed method outperforms them by large margins for all metrics tested. It seems, therefore, that explicit supervised learning of both tasks may be mutually beneficial and is particularly useful in estimating lesion age.

To better understand how the proposed model estimates lesion age, we produce saliency maps for test images by

TABLE III

LESION AGE ESTIMATION AND SEGMENTATION (MEAN $\pm$ STANDARD DEVIATION) RESULTS OBTAINED BY OUR METHOD AND ABLATION VARIANTS COMPARED TO THE SINGLE-TASK BASELINE MODELS

Model	Size	Flops	Regression			Classification		Segmentation		
			R^2	MAE	RMSE	AUC	ACC	DSC	IOU	LD-ACC
Intensity GLM	2	2	0.365	0.816	1.021	0.858	79.5	—	—	—
ResNet-50	24M	21G	0.308	0.862	1.115	0.906	83.1	—	—	—
ResNeXt-50	23M	22G	0.402	0.800	1.037	0.908	86.5	—	—	—
ConvNeXt-T	28M	23G	0.392	0.812	1.011	0.905	86.0	—	—	—
Ours	40M	30G	**0.513**	0.680	**0.935**	**0.933**	**88.5**	38.2 $\pm$ 24.2	**26.6 $\pm$ 21.0**	**98.0**
2D U-Net	8M	48G	—	—	—	—	—	35.3 $\pm$ 30.0	26.2 $\pm$ 26.2	95.3
3D U-Net	39M	91G†	—	—	—	—	—	36.7 $\pm$ 28.2	26.4 $\pm$ 26.4	97.3
TransUNet	108M	169G	—	—	—	—	—	36.9 $\pm$ 27.7	26.1 $\pm$ 25.9	97.3
(P-value)*								(0.038)	(0.049)	
ResNet-50 (L_1)	24M	21G	0.297	0.866	1.124	0.904	81.7	—	—	—
ResNeXt-50 (L_1)	23M	22G	0.396	0.809	1.112	0.907	84.5	—	—	—
ConvNeXt-T (L_1)	28M	23G	0.388	0.824	1.066	0.902	85.3	—	—	—
Ours (L_1)	40M	30G	0.503	**0.636**	0.944	0.912	86.5	**38.2 $\pm$ 24.1**	26.3 $\pm$ 21.1	98.0
Ours (only segmentation)	40M	30G	—	—	—	—	—	37.2 $\pm$ 25.1	25.3 $\pm$ 24.2	97.3
Ours (only age estimation)	40M	30G	0.510	0.682	0.938	0.930	86.8	—	—	—
Ours (no PPM)	30M	28G	0.330	0.733	1.097	0.874	79.7	36.0 $\pm$ 24.0	24.9 $\pm$ 20.8	96.6
Ours (no GPSA)	40M	30G	0.449	0.664	0.995	0.913	83.8	35.4 $\pm$ 24.6	24.9 $\pm$ 20.7	95.3
Ours (no RLW)	40M	30G	0.402	0.675	1.036	0.904	83.4	35.0 $\pm$ 25.0	24.5 $\pm$ 21.5	95.3
Ours (no DA)	40M	30G	0.025	0.945	1.357	0.756	71.6	31.6 $\pm$ 24.6	21.7 $\pm$ 20.6	94.6

MAE = Mean absolute error; RMSE = root mean squared error; AUC = Area under the receiver operator characteristic curve; ACC = Accuracy; DSC = Dice similarity coefficient; IOU = intersection over union; LD-ACC = Lesion detection accuracy; GLM = Generalized linear model; L_1 = L1 loss; PPM = Pyramid pooling module; GPSA = Gated positional self-attention; RLW = Random loss weighting; DA = Data augmentation

*P-values are between the results of the proposed model and the next best competing model
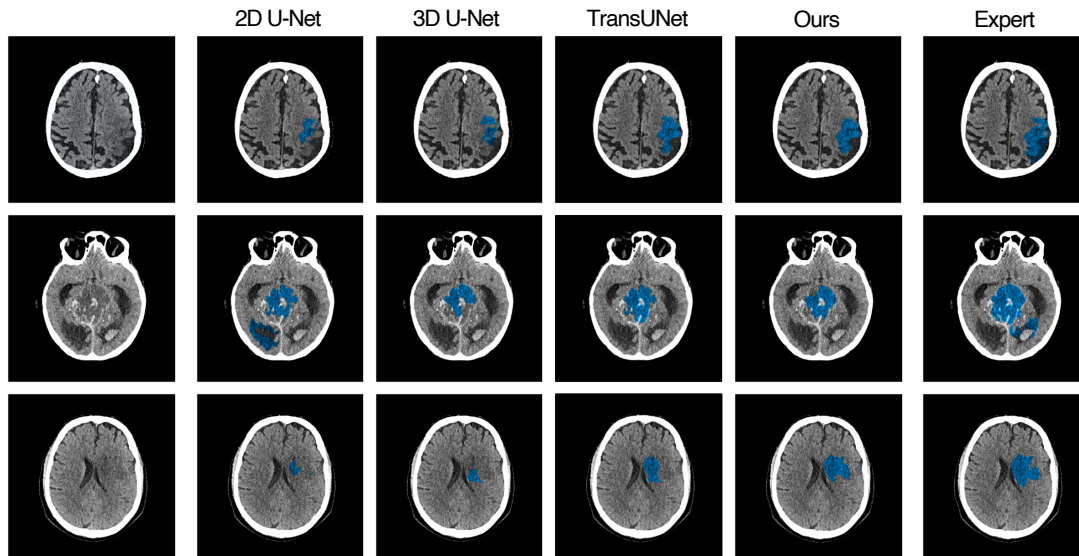†Normalized by total flops per subject divided by number of slices

Fig. 4. Example lesion segmentations of our method compared to the single-task baseline models.

backpropagating back to the input [50]. As seen in Fig. 5, compared to ResNet-50, the proposed method has generally more focused attention on the lesion. Interestingly, and similar to the aforementioned intensity approach, the proposed model also appears to be utilizing brain in the corresponding spatially distant unaffected side. This perhaps reinforces the benefit of the model's wide-receptive field afforded to it by the use of a Transformer.

2) Comparison with other Multi-Task Models: To further explore the synergy between tasks and the effectiveness of

the proposed method, we also compare it with recent multi-task learning networks shown in Table IV. These include 3D variants of MA-MTLN[51] and $C_{MS}VNet_{Iter}$[52], which are capable of joint segmentation and classification though not regression of medical images. For both lesion age estimation and segmentation, we find that the multi-task models perform equivalently or are superior to the single-task baseline models. We also note that the proposed method achieves the best performance of the multi-task models and has a greater lead in lesion classification than segmentation. It's possible that
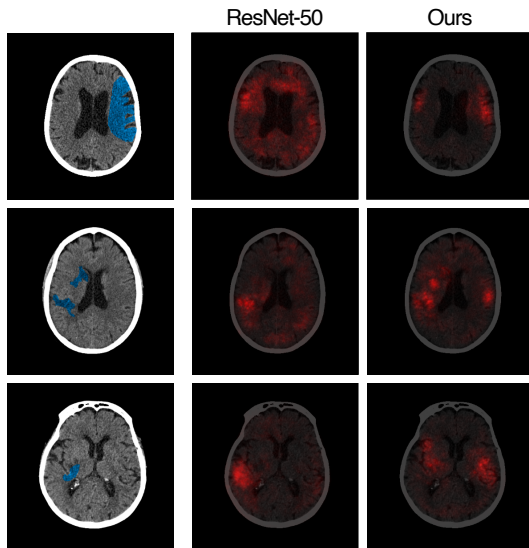
ResNet-50   Ours

Fig. 5. Saliency maps for estimating lesion age of the proposed method compared to ResNet-50. The proposed method has more focused attention on the lesion and appears to have learned to utilize brain from the unaffected corresponding side.

Fig. 7. Receiver operating characteristic (ROC) curves of our method compared to the best competing single- and multi-task baselines for classifying lesions ages.
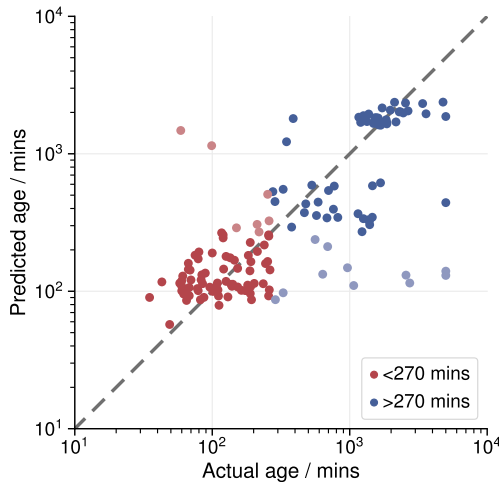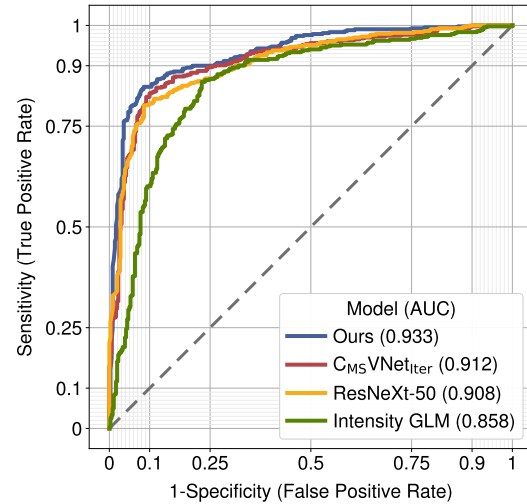

Fig. 6. Scatter plot of predicted versus actual lesion ages for the proposed model on the test set.

this disparity, favoring classification, may be the result of the proposed model being able to utilize continuous lesion ages, whereas the other models were architecturally limited to binary labels.

3) Comparison with the State-of-the-Art: There are few works that we can compare our results. For segmentation, we are aware of only two studies [11], [7] that used ground truth NCCT annotations. As argued by El-Hariri *et al.* [7], direct comparison with studies using annotations from other modalities such as MRI is hindered by the different underlying physiological processes which lead to visible changes. Compared with these studies, the proposed model performs slightly better on this challenging task with a DSC of 38.2% compared to 37% by Barros *et al.* [11] and 37.7% by El-Hairi *et al.* [7]. For lesion age estimation, the proposed model achieved an AUC of

0.933 for classifying whether a stroke event is within 4.5 hours of onset. Similar to the predominately manual methods by Broocks *et al.* [12] and Mair *et al.* [14] with reported AUC of 0.91 and 0.955, respectively. However, we note that due to the dynamic nature of ischemia, the classification of older lesions is considerably easier. This is noticeable in Fig. 6 where the proposed model predictions showed better agreement with lesions of a greater age. Therefore, the difficulty of this task is highly dependent on the distribution of lesion ages in the dataset, and without an open benchmark, objective assessment against other methods is limited. This is further supported by Fig. 7 where our intensity model achieves an AUC of only 0.858 using a similar methodology to these studies. Performance aside and in contrast to prior works, the proposed approach benefits from being fully automated and naturally able to accommodate patients with multiple lesions.

4) Generalization to Other Datasets: To the best of our knowledge, there are no publicly available datasets for ischemic lesion age estimation. For this reason, we utilized an additional independently collected dataset to further validate the performance of our approach with the results shown in Table V. Encouragingly, these findings reveal no unexpected outcomes, with our method performing best in both regression and classification tasks, which aligns with our initial analysis.

For lesion segmentation, while there are no datasets with ground truth NCCT annotations, the ISLES (Ischemic Stroke Lesion Segmentation) challenge [53] contains images with labels derived from diffusion-weighted imaging. As previously noted, this represents a related, although inherently different, task [7]. Additionally, the population characteristics are markedly different with lower resolution scans and younger patients with higher National Institutes of Health Stroke Scale (NIHSS) scores [53]. Therefore, we also assess out-of-domain generalization by comparing the performance of the trained proposed model to trained single- and multi-task models on

TABLE IV

LESION SEGMENTATION AND CLASSIFICATION RESULTS (MEAN ± STANDARD DEVIATION) OBTAINED BY OUR METHOD COMPARED TO MULTI-TASK MODELS

| Model | Size | Flops | Classification | | Segmentation | | |
			AUC	ACC	DSC	IOU	LD-ACC
MA-MTLN	10M	30G†	0.907	86.0	37.0 ± 23.2	26.4 ± 26.2	96.6
C_{MS}VNet$_{Iter}$	92M	95G†	0.912	86.5	37.4 ± 26.4	**26.7 ± 24.3**	97.3
Ours	40M	30G	**0.933**	**88.5**	**38.2 ± 24.2**	26.6 ± 21.0	**98.0**
(P-value)*					(0.047)	(0.449)	

AUC = Area under the receiver operator characteristic curve; ACC = Accuracy; DSC = Dice similarity coefficient; IOU = intersection over union; LD-ACC = Lesion detection accuracy

*P-values are between the results of the proposed model and the next best competing model

†Normalized by total flops per subject divided by number of slices

TABLE V

LESION AGE ESTIMATION RESULTS OBTAINED BY OUR METHOD COMPARED TO THE SINGLE- AND MULTI-TASK MODELS FOR A SECOND INDEPENDENT TEST SET

| Model | Regression | | | Classification | |
	R^2	MAE	RMSE	AUC	ACC
Intensity GLM	0.124	0.920	1.331	0.772	78.0
ResNet-50	0.280	0.861	1.102	0.892	83.3
ResNeXt-50	0.382	0.795	1.043	0.904	86.7
ConvNeXt-T	0.385	0.804	1.031	0.905	86.0
Ours	**0.511**	**0.682**	**0.958**	**0.921**	**88.7**
MA-MTLN	—	—	—	0.911	86.7
C_{MS}VNet$_{Iter}$	—	—	—	0.907	87.3

MAE = Mean absolute error; RMSE = root mean squared error; AUC = Area under the receiver operator characteristic curve; ACC = Accuracy

TABLE VI

LESION SEGMENTATION RESULTS (MEAN ± STANDARD DEVIATION) OBTAINED BY OUR METHOD COMPARED TO THE SINGLE- AND MULTI-TASK MODELS FOR THE ISLES-2018 DATASET

Model	DSC	IOU	LD-ACC
2D U-Net	18.3 ± 12.9	10.2 ± 9.3	73.0
3D U-Net	11.1 ± 13.7	6.5 ± 8.7	71.4
TransUNet	17.7 ± 14.6	9.8 ± 10.1	71.4
Ours	**20.3 ± 11.5**	11.2 ± 9.1	**74.6**
MA-MTLN	19.7 ± 14.1	**11.4 ± 9.6**	73.0
C_{MS}VNet$_{Iter}$	19.8 ± 13.8	11.0 ± 9.8	73.0
(P-value)*	(0.187)	(0.356)	

DSC = Dice similarity coefficient; IOU = intersection over union; LD-ACC = Lesion detection accuracy

*P-values are between the results of the proposed model and the next best competing model

5) Ablation Study: We conducted a series of experiments, shown in Table III, to verify the effectiveness of our method and justify its design decisions. First, we observe that our data augmentation strategy appears to have the largest impact on lesion age estimation and segmentation performance. It seems plausible this may be the result of overfitting from the limited data combined together with the strategy of training from scratch. Second, by jointly training age estimation and segmentation, the performance of both tasks appear to improve modestly. Third, using GPSA, PPM, and RLW rather than equally weighted losses provide benefits primarily to age estimation with comparatively little effect on segmentation. Finally, we note a consistent increase in lesion age estimation performance gained by using our proposed quantile loss based method across all tested models.

IV. DISCUSSION

Stroke is a leading cause of adult disability and death worldwide. Effective clinical management often relies upon the interpretation of CT imaging to confirm both the occurrence and age of an event. Previous attempts to automate these tasks have treated them independently and thereby may have overlooked their apparent complementary relationship. In the present study, a novel transformer-based approach is proposed to address this that is able to jointly segment and estimate the age of cerebral ischemic lesions. The performance of our method was then characterized through a number of experiments.

A. Limitations

It should be noted that this study had several limitations. As the methodology relied heavily on supervised machine learning it was, therefore, subject to many general issues and particularly those common to medical image analysis. First, and perhaps the most important, limited sample size due to the amount of data available publicly and the laborious nature of manually annotating additional subjects [55]. Second, difficulties in comparing algorithms in an objective manner due to the different populations and evaluation techniques of other published works [56]. Third, the introduction of potential bias through the use of data that may not represent the wider population and therefore hinder generalizability. This is of particular importance as all CT images used were obtained

the ISLES 2018 dataset, with results presented in Table VI.

The proposed method achieved the highest DSC and LDD-ACC scores of 20.3% and 74.6%, respectively. However, perhaps unsurprisingly, the performance of all models was considerably worse than the best-published approach [54] for this challenge, with a DSC of 51%. Giving further support to the motivation behind this work, we note that the multi-task models generally outperform the single-task models. Interestingly, it also seems apparent that the 2D models are able to generalize better than the 3D models, which may suggest they are more robust to the differences in slice thickness.

from Siemens scanners and previous studies have shown significant variation between scanners and manufacturers [57], [58]. There are also limitations specific to this work. Only one experienced scan reader was used to annotate the images. The labels for lesion age are likely subject to significant noise as they rely upon patients accurately reporting the time of symptom onset. Finally, the performance of the proposed model may be limited due to it's underlying 2D nature, as there are likely aspects that cannot be captured well due to its inability to leverage the whole 3D volume.

B. Future Work

Subsequent research could seek to address these aforementioned limitations. For example, by extending the proposed method to 3D, which appears relatively straightforward and can be achieved by replacing the backbone CNN encoder and segmentation head with 3D equivalents. However it seems likely that other changes may also be required to maintain computational efficiency. Additionally, before considering real-world deployment, the safety and applicability of our approach must be assessed prospectively in a larger and more extensive clinical study.

C. Broader Impact

If successfully validated, it is hoped that this work could lead to better outcomes for stroke patients due to faster diagnosis and better choice of treatment. Additionally, by exclusively using NCCT imaging, our method has the potential to be widely applicable, reducing health inequality in areas where medical experts are limited, such as in low- and middle-income countries [59]. The proposed approach could also find uses outside the intended application, anywhere that concurrent segmentation and regression may be seen as beneficial. For example, within healthcare, to detect a pneumothorax or effusion on chest X-ray and estimate their volumes [60], [61]. Or in other domains, such as detecting faces and estimating their age [62].

V. CONCLUSION

In this paper, we proposed a novel transformer-based network for concurrent ischemic lesion segmentation and age estimation of CT brain. By incorporating GPSA layers and using a modality-specific data augmentation strategy, we enhanced the data efficiency of our method. Furthermore, we improved lesion age estimation performance by better combining multiple predictions through the incorporation of uncertainty. Extensive experiments on a clinical dataset demonstrated the effectiveness of our method compared to conventional and task-specific algorithms. Future work includes further prospective clinical validation and exploring the extension of the model to 3D.

REFERENCES

[1] W. H. Organization, "Global health estimates," 12 2018. [Online]. Available: https://www.who.int/healthinfo/global_burden_disease/en/

[2] J. L. Saver, "Time is brain—quantified," *Stroke*, vol. 37, no. 1, pp. 263–266, 2006.

[3] W. Hacke, M. Kaste, E. Bluhmki, M. Brozman, A. Dávalos, D. Guidetti, V. Larrue, K. R. Lees, Z. Medeghri, T. Machnig *et al.*, "Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke," *New England journal of medicine*, vol. 359, no. 13, pp. 1317–1329, 2008.

[4] D. L. Rimmele and G. Thomalla, "Wake-up stroke: clinical characteristics, imaging findings, and treatment option–an update," *Frontiers in neurology*, vol. 5, p. 35, 2014.

[5] H. Ma, B. C. Campbell, M. W. Parsons, L. Churilov, C. R. Levi, C. Hsu, T. J. Kleinig, T. Wijeratne, S. Curtze, H. M. Dewey *et al.*, "Thrombolysis guided by perfusion imaging up to 9 hours after onset of stroke," *New England Journal of Medicine*, vol. 380, no. 19, pp. 1795–1803, 2019.

[6] G. Thomalla, C. Z. Simonsen, F. Boutitie, G. Andersen, Y. Berthezene, B. Cheng, B. Cheripelli, T.-H. Cho, F. Fazekas, J. Fiehler *et al.*, "Mri-guided thrombolysis for stroke with unknown time of onset," *New England Journal of Medicine*, vol. 379, no. 7, pp. 611–622, 2018.

[7] H. El-Hariri, L. A. S. M. Neto, P. Cimflova, F. Bala, R. Golan, A. Sojoudi, C. Duszynski, I. Elebute, S. H. Mousavi, W. Qiu *et al.*, "Evaluating nnu-net for early ischemic change segmentation on non-contrast computed tomography in patients with acute ischemic stroke," *Computers in biology and medicine*, p. 105033, 2021.

[8] W. Qiu, H. Kuang, E. Teleg, J. M. Ospel, S. I. Sohn, M. Almekhlafi, M. Goyal, M. D. Hill, A. M. Demchuk, and B. K. Menon, "Machine learning for detecting early infarction in acute stroke with non–contrast-enhanced ct," *Radiology*, vol. 294, no. 3, pp. 638–644, 2020.

[9] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[10] H. Kuang, B. K. Menon, S. I. Sohn, and W. Qiu, "Eis-net: Segmenting early infarct and scoring aspects simultaneously on non-contrast ct of patients with acute ischemic stroke," *Medical Image Analysis*, vol. 70, p. 101984, 2021.

[11] R. S. Barros, W. E. van der Steen, A. M. Boers, I. Zijlstra, R. van den Berg, W. El Youssoufi, A. Urwald, D. Verbaan, P. Vandertop, C. Majoie *et al.*, "Automated segmentation of subarachnoid hemorrhages with convolutional neural networks," *Informatics in Medicine Unlocked*, vol. 19, p. 100321, 2020.

[12] G. Broocks, H. Leischner, U. Hanning, F. Flottmann, T. D. Faizy, G. Schön, P. Sporns, G. Thomalla, S. Kamalian, M. H. Lev *et al.*, "Lesion age imaging in acute stroke: water uptake in ct versus dwi-flair mismatch," *Annals of Neurology*, vol. 88, no. 6, pp. 1144–1152, 2020.

[13] J. Minnerup, G. Broocks, J. Kalkoffen, S. Langner, M. Knauth, M. N. Psychogios, H. Wersching, A. Teuber, W. Heindel, B. Eckert *et al.*, "Computed tomography–based quantification of lesion water uptake identifies patients within 4.5 hours of stroke onset: A multicenter observational study," *Annals of neurology*, vol. 80, no. 6, pp. 924–934, 2016.

[14] G. Mair, A. Alzahrani, R. I. Lindley, P. A. Sandercock, and J. M. Wardlaw, "Feasibility and diagnostic accuracy of using brain attenuation changes on ct to estimate time of ischemic stroke onset," *Neuroradiology*, vol. 63, no. 6, pp. 869–878, 2021.

[15] S. d'Ascoli, H. Touvron, M. L. Leavitt, A. S. Morcos, G. Biroli, and L. Sagun, "Convit: Improving vision transformers with soft convolutional inductive biases," in *International Conference on Machine Learning*. PMLR, 2021, pp. 2286–2296.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[17] S. Vandenhende, S. Georgoulis, M. Proesmans, D. Dai, and L. Van Gool, "Revisiting multi-task learning in the deep learning era," *arXiv preprint arXiv:2004.13379*, vol. 2, no. 3, 2020.

[18] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[19] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[21] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*. Springer, 2020, pp. 213–229.

[22] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[23] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," *Advances in neural information processing systems*, vol. 30, 2017.

[24] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

[26] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[28] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[29] J.-B. Cordonnier, A. Loukas, and M. Jaggi, "On the relationship between self-attention and convolutional layers," *arXiv preprint arXiv:1911.03584*, 2019.

[30] O. Vinyals, S. Bengio, and M. Kudlur, "Order matters: Sequence to sequence for sets," *arXiv preprint arXiv:1511.06391*, 2015.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[32] Z. Zhou, V. Sodha, M. M. Rahman Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang, "Models genesis: Generic autodidactic models for 3d medical image analysis," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 384–393.

[33] P. W. Wong and C. Herley, "Area based interpolation for image scaling," Mar. 30 1999, uS Patent 5,889,895.

[34] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[35] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658–666.

[36] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[37] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[38] B. Lin, F. Ye, and Y. Zhang, "A closer look at loss weighting in multi-task learning," *arXiv preprint arXiv:2111.10603*, 2021.

[39] J. Muschelli, "Recommendations for processing head ct data," *Frontiers in neuroinformatics*, vol. 13, p. 61, 2019.

[40] P. A. Barber, A. M. Demchuk, J. Zhang, A. M. Buchan, A. S. Group *et al.*, "Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy," *The Lancet*, vol. 355, no. 9216, pp. 1670–1674, 2000.

[41] T. Brott, H. P. Adams Jr, C. P. Olinger, J. R. Marler, W. G. Barsan, J. Biller, J. Spilker, R. Holleran, R. Eberle, and V. Hertzberg, "Measurements of acute cerebral infarction: a clinical examination scale." *Stroke*, vol. 20, no. 7, pp. 864–870, 1989.

[42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[43] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[44] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2017, pp. 464–472.

[45] T. Mikolov, "Statistical language models based on neural networks," Ph.D. dissertation, Brno University of Technology, 2012.

[46] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, "Averaging weights leads to wider optima and better generalization," *arXiv preprint arXiv:1803.05407*, 2018.

[47] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[48] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[49] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *arXiv preprint arXiv:2201.03545*, 2022.

[50] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[51] Y. Zhang, H. Li, J. Du, J. Qin, T. Wang, Y. Chen, B. Liu, W. Gao, G. Ma, and B. Lei, "3d multi-attention guided multi-task learning network for automatic gastric tumor segmentation and lymph node classification," *IEEE Transactions on Medical Imaging*, vol. 40, no. 6, pp. 1618–1631, 2021.

[52] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, P.-T. Yap, and D. Shen, "Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images," *Medical Image Analysis*, vol. 70, p. 101918, 2021.

[53] A. Hakim, S. Christensen, S. Winzeck, M. G. Lansberg, M. W. Parsons, C. Lucas, D. Robben, R. Wiest, M. Reyes, and G. Zaharchuk, "Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: lessons from the isles challenge," *Stroke*, vol. 52, no. 7, pp. 2328–2337, 2021.

[54] T. Song, "3d multi-scale u-net with atrous convolution for ischemic stroke lesion segmentation," *Proc. MICCAI ISLES*, 2018.

[55] I. Balki, A. Amirabadi, J. Levman, A. L. Martel, Z. Emersic, B. Meden, A. Garcia-Pedrero, S. C. Ramirez, D. Kong, A. R. Moody *et al.*, "Sample-size determination methodologies for machine learning in medical imaging research: a systematic review," *Canadian Association of Radiologists Journal*, vol. 70, no. 4, pp. 344–353, 2019.

[56] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC medicine*, vol. 17, no. 1, p. 195, 2019.

[57] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany *et al.*, "Reliability of mri-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer," *Neuroimage*, vol. 32, no. 1, pp. 180–194, 2006.

[58] H. Takao, N. Hayashi, and K. Ohtomo, "Effects of study design in multi-scanner voxel-based morphometry studies," *Neuroimage*, vol. 84, pp. 133–140, 2014.

[59] B. Wahl, A. Cossy-Gantner, S. Germann, and N. R. Schwalbe, "Artificial intelligence (ai) and global health: how can ai contribute to health in resource-poor settings?" *BMJ global health*, vol. 3, no. 4, p. e000798, 2018.

[60] K. Hoi, B. Turchin, and A.-M. Kelly, "How accurate is the light index for estimating pneumothorax size?" *Australasian radiology*, vol. 51, no. 2, pp. 196–198, 2007.

[61] C. Brockelsby, M. Ahmed, and M. Gautam, "P1 pleural effusion size estimation: Us, cxr or ct?" *Thorax*, vol. 71, no. Suppl 3, pp. A83–A83, 2016. [Online]. Available: https://thorax.bmj.com/content/71/Suppl_3/A83.1

[62] A. Othmani, A. R. Taleb, H. Abdelkawy, and A. Hadid, "Age estimation from faces using deep learning: A comparative analysis," *Computer Vision and Image Understanding*, vol. 196, p. 102961, 2020.