Original Article

A convolutional neural network combined with positional and textural attention for the fully automatic delineation of primary nasopharyngeal carcinoma on non-contrast-enhanced MRI

Lun M. Wong[1], Qi Yong H. Ai[1], Darren M. C. Poon[2], Macy Tong[2], Brigette B. Y. Ma[2], Edwin P. Hui[2], Lin Shi[1], Ann D. King[1]

[1]Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China; [2]Department of Clinical Oncology, State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China

Contributions: (I) Conception and design: LM Wong, QYH Ai, AD King; (II) Administrative support: QYH Ai, AD King; (III) Provision of study materials or patients: QYH Ai, DMC Poon, M Tong, BBY Ma, EP Hui, AD King; (IV) Collection and assembly of data: QYH Ai, DMC Poon, M Tong, BBY Ma, EP Hui, AD King; (V) Data analysis and interpretation: LM Wong, QYH Ai, AD King; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Prof. Ann D. King. Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, New Territories, Hong Kong, China. Email: king2015@cuhk.edu.hk.

Background: Convolutional neural networks (CNNs) have the potential to automatically delineate primary nasopharyngeal carcinoma (NPC) on magnetic resonance imaging (MRI), but currently, the literature lacks a module to introduce valuable pre-computed features into a CNN. In addition, most CNNs for primary NPC delineation have focused on contrast-enhanced MRI. To enable the use of CNNs in clinical applications where it would be desirable to avoid contrast agents, such as cancer screening or intra-treatment monitoring, we aim to develop a CNN algorithm with a positional-textural fully-connected attention (FCA) module that can automatically delineate primary NPCs on contrast-free MRI.

Methods: This retrospective study was performed in 404 patients with NPC who had undergone staging MRI. A proposed CNN algorithm incorporated with our positional-textural FCA module ($A_{proposed}$) was trained on manually delineated tumours (M_{1st}) to automatically delineate primary NPCs on non-contrast-enhanced T2-weighted fat-suppressed (NE-T2W-FS) images. The performance of $A_{proposed}$, three well-established CNNs, Unet (A_{unet}), Attention-Unet (A_{att}) and Dense-Unet (A_{dense}), and a second manual delineation repeated to evaluate human variability (M_{2nd}) were measured by comparing to the reference standard M_{1st} to obtain the Dice similarity coefficient (DSC) and average surface distance (ASD). The Wilcoxon rank test was used to compare the performance of $A_{proposed}$ against A_{unet}, A_{att} , A_{dense} and M_{2nd}.

Results: $A_{proposed}$ showed a median DSC of 0.79 (0.10) and ASD of 0.66 (0.84) mm. It performed better than the well-established networks A_{unet} [DSC =0.75 (0.12) and ASD =1.22 (1.73) mm], A_{att} [DSC =0.75 (0.10) and ASD =0.96 (1.16) mm] and A_{dense} [DSC =0.71 (0.14) and ASD =1.67 (1.92) mm] (all P<0.01), but slightly worse when compared to M_{2nd} [DSC =0.81 (0.07) and ASD =0.56 (0.80) mm] (P<0.001).

Conclusions: The proposed CNN algorithm has potential to accurately delineate primary NPCs on non-contrast-enhanced MRI.

Keywords: Texture; convolutional neural network (CNN); nasopharyngeal carcinomas (NPCs); head and neck; magnetic resonance imaging (MRI)

Submitted Feb 18, 2021. Accepted for publication May 13, 2021.
doi: 10.21037/qims-21-196
View this article at: http://dx.doi.org/10.21037/qims-21-196

Introduction

Convolutional neural networks (CNNs) are machine learning techniques which exploit serial stacks of trainable convolutional image filters and non-linear activation layers for data modelling. Recently, they have been used to rapidly automate a wide-range of radiological tasks (1,2). Cancer delineation is one of the imaging applications in which CNN performs well. The technique shows promises in automating the laborious and time-consuming task of manually delineating cancer margins on serial sections, which is required for cancer management purposes such as tumour detection, the prediction of outcomes and treatment planning.

Current CNN-based automatic tissue delineation research focuses on making modifications to well-established CNN architectures such as the Unet to delineate different tissues of interest (3). However, CNNs have intrinsic limitations inherited from the convolutional operations. CNNs cannot mathematically replicate some textural features, such as those from the grey-level co-occurrence matrix, that are known to be useful in image classification (4,5). Furthermore, CNNs performs poorly at retaining or extracting positional information from intensity maps (6), an attribute that is especially important in a patch-based setting where the CNN does not have access to the position of the cropped patches relative to the original image. However, a unified module to introduce these features into a CNN is still lacking in the literature. Therefore, we propose a fully-connected attention (FCA) module that incorporates both textural and 3D positional information computed prior to training, and employed it in a patch-based CNN designed based on the Attention-Unet (7).

To evaluate the performance, we applied our proposed CNN algorithm to delineate primary nasopharyngeal carcinoma (NPC) on magnetic resonance imaging (MRI). This is one of the most challenging cancers to delineate on MRI because of the highly-complex anatomy of the nasopharynx and the surrounding structures at the skull base. In this study, we compared our CNN algorithm with well-established networks: Unet (3), Attention-Unet (7) and 2D Dense-Unet (8). In addition, unlike the previously reported MRI studies of primary NPC delineation by CNNs in the literature (9-14), this study evaluated the delineation performance on the non-contrast-enhanced T2-weighted fat-suppressed (NE-T2W-FS) sequence rather than the contrast-enhanced T1-weighted sequence. This sequence was chosen to support our ongoing research into early NPC detection by MRI (15-17) in which we are developing a low-cost MRI protocol for NPC Epstein-Barr virus DNA screening programs that does not require the injection of an MRI contrast agent (18,19). Our previous study has also shown the NE-T2W-FS sequence is promising for primary NPC delineation with CNN (20), but the general-purpose CNN tested in that study would benefit from customisations. In addition, NE-T2W-FS has other potential applications in head and neck cancer imaging where an MRI contrast agent cannot be administered, such as in patients with renal failure (21). Furthermore, because gadolinium-based contrast agents have recently been shown to accumulate in the body, there is greater caution in the radiology community concerning the use of these agents (22) and a greater move towards using non-contrast-enhanced sequences in repeated scans, such as intra-treatment response monitoring and surveillance imaging.

With these issues in mind, in this study, we propose a CNN algorithm that combines 3D position and 2D texture information to automatically delineate primary NPCs on non-contrast-enhanced T2-weighted MRI. We evaluate the performance of the proposed algorithm for primary NPC delineation and compare the performance to that of human experts using manual delineation and well-established 2D delineation networks.

Methods

Patients characteristics

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (Approval ID: CIE-2019.709), requirements of written consents were waived owing to its retrospective nature. This study reviewed 453 patients with newly diagnosed histologically proven NPC who underwent head and neck staging MRI between January 2010 and May 2015 retrospectively. Patients with the following criteria were excluded: (I) incomplete or inconsistent MRI protocols (n=26) and (II) MRI severely degraded by artefact or movement (n=23). This left 404 patients for analysis. All primary tumours were staged according to the 8th edition of the American Joint Committee on Cancer staging manual (23).

Data acquisition

MRI was performed with a Philips Achieva TX 3.0-T machine (Philips Healthcare, Best, the Netherlands) using a body coil for radiofrequency transmission and a 16-channel Philips neurovascular phased-array coil for reception. The MRI sequences for CNN segmentation used an axial NE-T2W-FS sequence [repetition time/echo time, 4,000/80 ms; field of view (FOV), 230×230 mm; section thickness, 4 mm; echo train length, 15–17; sensitivity encoding factor, 1; number of signals acquired, 2]. The FOVs of the scans were centred approximately at the posterior wall of the nasopharynx. The final dimension of each axial images was 512×512 pixels, with a pixel size of 0.45×0.45 mm. The images were normalised using the Z-score normalisation technique.

Manual delineation for primary tumour

All primary NPCs were manually delineated on the axial NE-T2W-FS images (M_{1st}) with references to all anatomical MRI including contrast-enhanced images by an expert with more than 6 years of experience in NPC using the open-source software ITK-SNAP v3.2.0 (24). The M_{1st} was used to train the CNN and as the reference standard with which to evaluate performance.

To assess human variability, a second set of primary NPC manual delineations (M_{2nd}) was performed by the same observer at a time interval of at least 15 days.

Algorithm architecture

The proposed algorithm comprises four crucial components: the (I) discriminative patch sampling technique, (II) reflection-padding, (III) textural-positional FCA module and (IV) the CNN. Discriminative patch sampling favours high intensities when selecting the patch, effectively minimising the probability of sampling trivial empty patches and increases the probability of sampling hyperintense tumour regions (Appendix 1). Reflection-padding mitigates local contrast sharpening at the image edges and compensates for the shrinkage of the receptive field caused by the convolutional layers (Appendix 1). The proposed textural-positional FCA module was built on two textural filters, the local binary pattern (LBP) (25) and local neighbourhood differences pattern (LNDP) (26), which were embedded together with the positional information into the CNN (Appendix 2). The network architecture was based on the Attention-Unet (7), which was adapted from the Unet (3) with additional convolutional attention layers to capture semantic information. In this study, we replaced the attention layers with the FCA module to introduce textural and positional information. The proposed algorithm automatically computes the pixel-wise probability of the presence of tumour. The details of the CNN design architecture are shown in *Figure 1* and the details of these modules are described in Appendix 1.

Algorithm training and validation

The proposed CNN was implemented and trained using M_{1st} as the reference standard with PyTorch (27) to obtain a set of automatic CNN delineations defined as $A_{proposed}$. We performed random data augmentations over all training data to improve network generality and robustness (28). The key training parameters are shown in *Table 1*. Further details on the data augmentations are shown in Appendix 3.

Four-fold cross-validation was performed to validate the performance of the algorithm. The dataset was randomly divided into four even partitions, each with 101 individual cases. The patient characteristics of the cohorts are plotted in *Table 2*. Additional technical details are reported in Appendix 4.

We obtained additional sets of automatic tumour delineations using three well-established 2D delineation CNNs: Unet (3) (A_{unet}), Attention-Unet (A_{att}) (7) and Dense-Unet-167 (A_{dense}) (8) with identical training configurations on the first fold of the data. All training parameters followed the values listed in *Table 1* except for the mini-batch size, which was tuned according to the memory requirements of the network. The training of these networks was performed with identical patch-based settings to our proposed network and they were all configured to have identically five encoder-decoder convolutional levels, like ours.

Performance evaluation

M_{1st} was used as the reference standard to evaluate the performance of the automatic CNN delineations ($A_{proposed}$, A_{unet}, A_{att}, and A_{dense}) and to assess variations in the manual delineation M_{2nd}. The performance metrics were the Dice similarity coefficient (DSC), correspondence ratio (CR) and percentage match (PM), which measure the volumetric agreement between the compared delineations, and dissimilarity metric average surface distance (ASD), which measures the boundaries differences of the compared delineations.

The definitions of these indices are listed below:

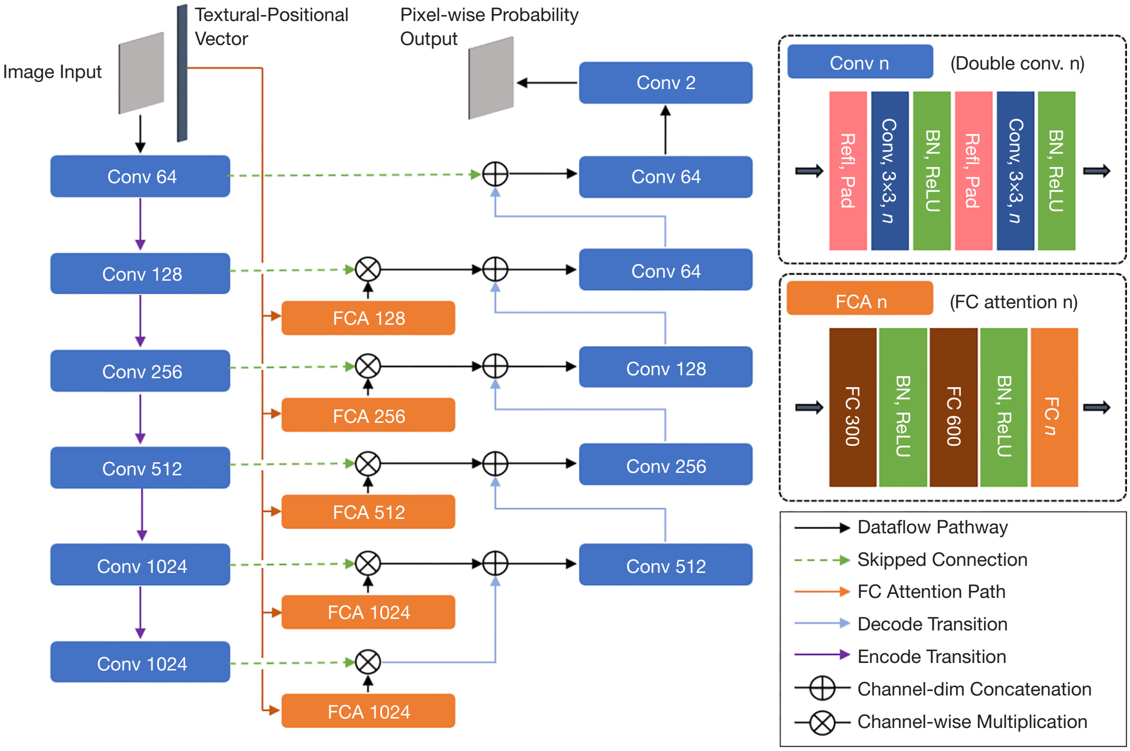

Figure 1 Detailed architecture of the designed network. The network consists of five convolutional levels, each of the levels is composed of an encoder block, a decoder block and an FCA block. Each encoder transition performs maximum pooling to subsample the input by a factor of 2 while each decoder transition interpolates the input by a factor of 2. The number following the block name indicates how many channels the block output possesses. All of the convolutional layers have a kernel size of 3×3 and the inputs are padded with reflection padding before each convolution. The definitions of individual blocks are given in the dotted boxes. The network receives image patch input together with the textural-positional vector computed from corresponding patches. The output of the network is a tensor with two channels representing the pixel-wise probability of tumour absence and presence. FCA, fully-connected attention layer; Conv, convolutional layer; Refl, reflection; ReLU, rectifying linear unit; BN, batch-norm layer.

Table 1 Key training parameters

Training parameters	Values
Initial learning rate	1×10^{-4}
Training mini-batch size	45
Learning rate decay	0.005
Total epochs ran	150

$$\mathrm{DSC}=\frac{2TP}{2TP+FP+FN} \quad [1]$$

$$\mathrm{CR}=\frac{2TP-FP}{2(TP+FN)} \quad [2]$$

$$\mathrm{PM}=\frac{TP}{TP+FN} \quad [3]$$

$$\mathrm{ASD}=\frac{1}{n_g+n_t}\left(\sum_{i=1}^{n_g}\left\|g_i-t_i'\right\|_2+\sum_{i=1}^{n_t}\left\|t_i-g_i'\right\|_2\right) \quad [4]$$

where *TP*, *TN*, *FP* and *FN* are conventional abbreviations of true positive, true negative, false positive and false negative pixel counts respectively; n_g and n_t denotes total number of surface elements g_i and t_i in the ground-truth label and the tested label respectively. The primed variable t_i' denotes the surface element in the tested label with smallest distance to the i-th element in the ground-truth label g_i, and vice versa for g_i'.

Table 2 Patient characteristics

	Entire cohort (n=404)	Fold 1 (n=101)	Fold 2 (n=101)	Fold 3 (n=101)	Fold 4 (n=101)	P values
Age (years)	53.5, 19.0–90.0	51.8, 25.0–79.0	53.4, 31.0–90.0	54.3, 19.0–83.0	54.5, 27.0–81.0	0.314
Sex						0.555
Man	313	75	76	79	83	–
Woman	91	26	25	22	18	–
Primary T classification						0.512
T1	130	32	32	36	30	–
T2	59	14	14	17	14	–
T3	140	28	37	33	42	–
T4	75	27	18	15	15	–
Gross tumour volume (cm^3)	21.3±21.5, 1.4–134.5	24.3±24.8, 1.4–134.5	22.0±21.9, 2.0–117.4	19.1±18.7, 1.6–90.6	19.7±19.8, 2.6–124.2	0.330
Pathological classification						–
Undifferentiated carcinoma	391	97	98	99	97	–
Poorly differentiated carcinoma	13	4	3	2	4	–

Statistical analysis

The patient characteristics of the cohorts forming the four folds were tested by analysis of variance for any differences in age, sex, and stage distribution. The performance metrics of $A_{proposed}$ of the four folds were analysed using the Kruskal-Wallis test for any differences.

To evaluate the performance and robustness of the proposed CNN for tumour delineation with respect to the existing CNN techniques, the performance metrics of $A_{proposed}$ were compared to those of the well-established CNN delineations (A_{unet}, A_{att} and A_{dense}) obtained from fold 1 using a non-parametric paired-sample *t*-test (Wilcoxon rank test).

To compare the performance and robustness of the proposed CNN for tumour delineation with respect to the human expert, the performance metrics of the CNN delineations $A_{proposed}$ and M_{2nd} were compared using a non-parametric paired-sample *t*-test. The DSC, CR and PM of $A_{proposed}$ and M_{2nd} were plotted jointly together with the kernel density estimation (KDE), which estimated the joint probability density function between them. The DSC and PM were metrics confined by the range 0 to 1, whereas that of CR was $-\infty$ to 1, of which 1 indicates perfect agreement of $A_{proposed}$ or M_{2nd} to the referenced standard M_{1st}.

To investigate the influence of tumour stage (T-stage) on the performance of the proposed CNN delineation $A_{proposed}$ and manual delineation M_{2nd}, differences across the T-stages (T1–T4) were analysed using the Kruskal-Wallis test.

All of the statistical analyses were performed with SPSS v24 (IBM, Netherland). Statistical significance was accepted at P<0.05.

Results

Patient characteristics

The characteristics of the 404 patients are shown in *Table 2*. There were no statistically significant differences in patient characteristics across the four folds (all P>0.05) (*Table 2*).

Performance of the proposed compared to well-established CNNs

The median performance of $A_{proposed}$, A_{unet}, A_{att} and A_{dense} are shown in *Table 3*. For $A_{proposed}$, there were no statistically significant differences in the performance of the metrics across the four folds (P=0.657, 0.525, 0.177 and 0.571 for DSC, CR, PM and ASD respectively).

When compared with the three other CNNs tested on fold 1, $A_{proposed}$ showed better performance in all metrics (all P<0.001) (*Table 3*).

Table 3 Median of performance metrics in 4-fold cross-validation using first set of manual delineation M_{1st} as referenced standard

Methods/comparison	N	DSC	CR	PM	ASD (mm)
$A_{proposed}$					
All folds	404	0.79 (0.10)	0.69 (0.15)	0.83 (0.17)	0.66 (0.84)
Fold 1	101	0.79 (0.13)	0.69 (0.20)	0.80 (0.17)	0.83 (1.22)
Fold 2	101	0.79 (0.08)	0.70 (0.13)	0.84 (0.15)	0.60 (0.78)
Fold 3	101	0.80 (0.08)	0.71 (0.14)	0.84 (0.17)	0.63 (0.87)
Fold 4	101	0.79 (0.11)	0.68 (0.16)	0.84 (0.17)	0.65 (0.78)
A_{unet} (3)	101	0.75 (0.12)*	0.61 (0.18)*	0.74 (0.19)*	1.22 (1.73)*
A_{att} (7)	101	0.75 (0.10)*	0.65 (0.15)*	0.85 (0.19)*	0.96 (1.16)*
A_{dense} (8)	101	0.71 (0.14)*	0.59 (0.25)*	0.92 (0.12)*	1.67 (1.92)*
M_{2nd}	404	0.81 (0.07)*	0.71 (0.10)*	0.81 (0.12)*	0.56 (0.80)*

Data displayed are median (IQR). *, marks significant difference of from paired *t*-test with the proposed method (P<0.05). DSC, Dice similarity coefficient; CR, correspondence ratio; PM, percentage match; ASD, average surface distance; M_{1st}, manually drawn delineations set used as reference standard; M_{2nd}, 2nd manually drawn delineations set for intra-observer variability measurements; $A_{proposed}$, proposed CNN delineation; A_{unet}, delineations generated with Unet; A_{att}, delineations generated with Attention-Unet; A_{dense}, delineations generated with Dense-Unet-167.

Performance of $A_{proposed}$ compared to variability in human M_{2nd}

When the performance metrics of $A_{proposed}$ (*Table 3*) were compared to those obtained from the second manual delineation by the same observer M_{2nd} (*Table 3*), the metrics for CNN were slightly worse (all P<0.001). However, the KDE of performance metrics (*Figure 2*) suggested that the performance distributions of $A_{proposed}$ and M_{2nd} were similar with the KDE peak close to the line with a slope of 1. *Figure 3A,B* show two primary NPCs delineated by our proposed CNN algorithm with close agreement to with the manual delineation by the expert. *Figure 4A,B* show two cases with disagreement between CNN and manual delineation by the expert.

Influence of T-stage on performance

The performance metrics of $A_{proposed}$ showed no differences across T-stages for DSC and CR but showed significant differences for PM and ASD, with worse performance for PM and ASD with increasing T-stage (*Table 4*). The performance of M_{2nd} showed no differences in DSC and CR, but showed significant differences for PM and ASD, with worse performance for ASD with increasing T-stage (*Table 4*).

Discussion

Performance of the proposed CNN compared to the human expert

We proposed and tested a CNN algorithm that incorporates a textural-positional FCA module to delineate primary NPCs on a T2-weighted sequence. The T2-weighted sequence is unable to replace a contrast-enhanced MRI in clinical scenarios, such as radiotherapy planning, but it is an important sequence in MRI protocols that should not be overlooked in circumstances where it is desirable to avoid contrast, including NPC screening programs. Delineations of our proposed automatic CNN algorithm $A_{proposed}$ achieved a median DSC of 0.79 which was slightly lower than that of the second manual delineation M_{2nd} (median DSC of 0.81, P<0.001). However, substantial agreement was observed on the KDE plots where the datapoints were densely situated in the proximity of the line with a slope equal to 1. This plot suggests that our proposed CNN and the second manual delineation (i.e., reflecting variations that are observed when the expert repeats the delineation) have a high probability of obtaining similar DSC scores. This result is very encouraging for NPC screening given that our expert had the advantage of using information from all MRI sequences including the contrast-enhanced sequences

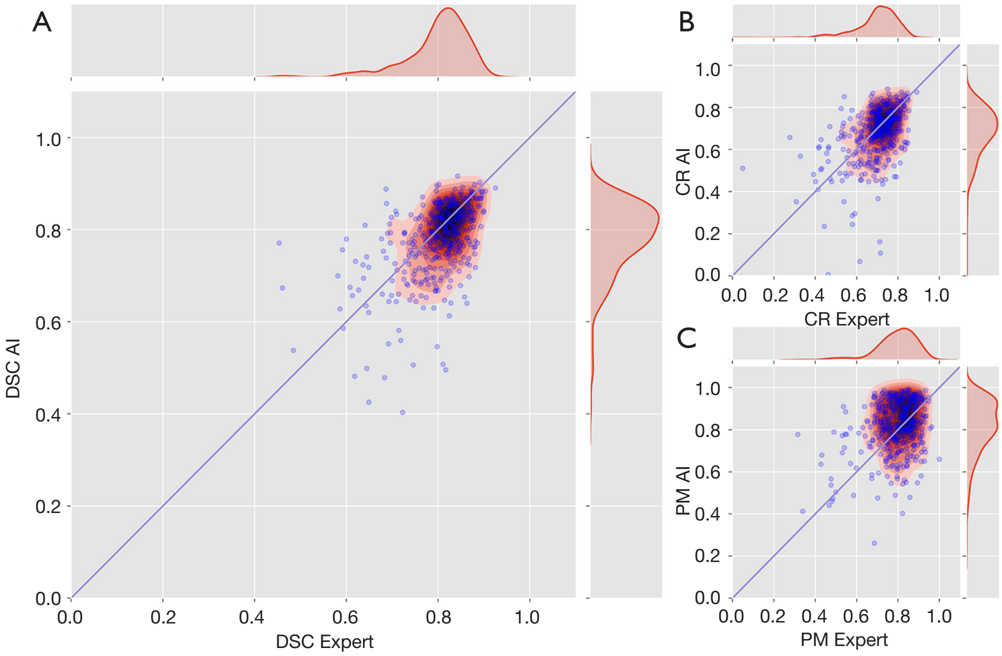

Figure 2 KDE of the joint probability density function of the proposed CNN and human expert performance. This figure provides an overview and comparison of how the CNN and human will perform on the same set of cases. Three sets of contours were involved: (I) ground-truth delineated by an expert used for both training and evaluation of performance M_{1st}, (II) delineation by the proposed CNN $A_{proposed}$ and (III) delineation by the same expert at least 15 days apart from the first set to measure intra-observer variability M_{2nd}. Three quantitative indices namely (A) DSC, (B) CR and (C) PM, were evaluated for $A_{proposed}$ and M_{2nd} against the referenced ground-truth M_{1st}. All DSC and PM are confined by the range 0 to 1, while CR has no negative value bound ($-\infty$ to 1). Individual cases are marked with scatter plots of blue dots. It is shown that the joint probability between CNN and human expert performance appreciably peaks at the proximity of the line with a slope of 1 for DSC and CR, suggesting a high probability of comparable performance in terms of these metrics. The KDE shows the CNN exhibits better PM as seen from the peak being situated above the blue line, this also matches the quantitative analysis. KDE, kernel density estimation; CNN, convolutional neural network; DSC, Dice similarity coefficient; CR, corresponding ratio; PM, percentage match.

and other scanning planes for the manual delineation, whereas our CNN algorithm had access only to the non-contrast-enhanced axial T2-weighted images. Furthermore, primary NPC is a very challenging target to delineate and it is recognised that a perfect DSC score may not reflect robustness and consistency. Mattiucci *et al.* (29) concluded that a mean DSC of 0.80, close to our results, can be considered as a good agreement for automatic contour generation in head and neck tumours.

Whereas DSC, CR and PM, evaluate the agreement in tumour volume overlap, ASD evaluates the accuracy of the tumour boundaries. In this study, the ASD of 0.66 mm in the CNN delineation $A_{proposed}$ was worse than that of the second manual delineation M_{2nd}, which had an ASD of 0.56 mm (P<0.001). However, this ASD value suggested that the CNN algorithm still predicted the margins with a median error of only <1 mm.

Influence of T-stage on delineation performance

We further investigated the delineation performance of the proposed CNN algorithm $A_{proposed}$ and M_{2nd} across T-stages. Our results showed that in both cases the T-stage influenced PM and ASD but not DSC or CR. The decrease in ASD performance with higher T-stage could be explained by the irregularly-shaped infiltrating margins that are associated with more locally advanced tumours, leading to greater variation in both human and machine delineation. The decrease in $A_{proposed}$ PM for advanced tumours is likely a result of an increased proportion of voxels that are tumour

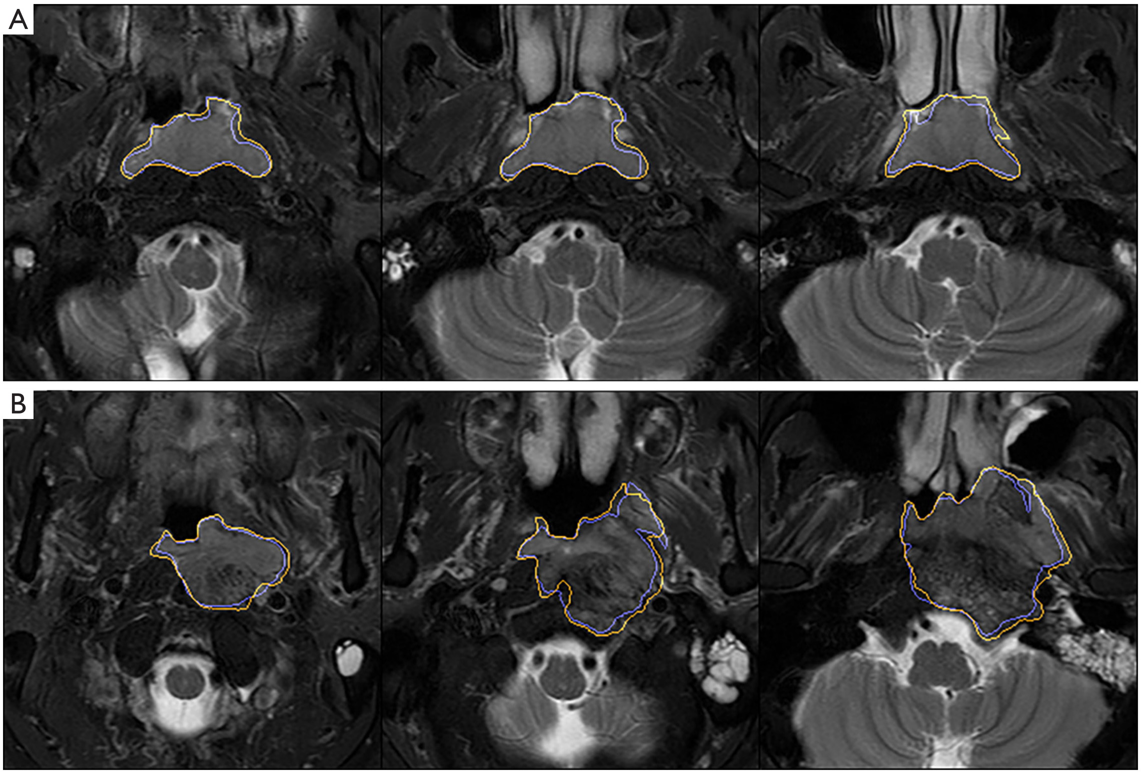

Figure 3 Primary NPC delineation overlaying the T2W-FS images of (A) a stage T1 tumour and (B) a stage T4 tumour. The automatic CNN delineation $A_{proposed}$ (yellow) closely overlaps the first manual delineation M_{1st} (purple), which was used as a reference standard for CNN training. In both early stage T1 and advanced stage T4 NPC cases, $A_{proposed}$ performed well even though it only had access to the T2W-FS images whereas expert delineation M_{1st} were delineated with referenced to all available MRI. Using M_{1st} as the reference standard, the DSC of $A_{proposed}$ in (A) was 0.87 and in (B) was 0.87. T2W-FS, T2-weighted fat-suppressed; CNN, convolutional neural network; NPC, nasopharyngeal carcinomas; DSC, Dice similarity coefficient.

positive and a reduction in the proportion that are negative, resulting in lower likelihoods of a false positive and higher likelihoods of a false negative.

Performance of $A_{proposed}$ compared to three well-established CNNs

We compared the performance of our proposed CNN which incorporated the textural-positional FCA module with Unet, Attention-Unet and 2D Dense-Unet-167, using fold 1 of our dataset.

Our delineation from the proposed CNN $A_{proposed}$ performed better than that from all of three well-established CNNs in delineating the primary NPCs (all P<0.05). Although the Unet is the most basic CNN amongst the three well-established CNNs tested, A_{unet} (DSC =0.75) performed similarly to A_{att} (DSC =0.75) and better than A_{dense} (DSC =0.71). The addition of attention modules to the Unet to form the Attention-Unet improved pancreatic tumour delineation (7), but did not improve primary NPC delineation in this study. The 2D Dense-Unet-167 introduced the dense-connection block to the Unet but resulted in worse delineation performance, potentially because it was not adapted to the patch-based setting used in this study. In our algorithm, we replaced the attention modules in Attention-Unet with our FCA module showing that incorporating both the texture features and 3D positions of the extracted patches in a patch-based delineation CNN setting allowed our $A_{proposed}$ to attain a significantly better performance and meet the challenge of delineating this complex-shaped cancer.

It should be noted that the 2D Dense-Unet-167 was extracted from the H-Dense-Unet (8). For liver lesions, the H-Dense-Unet has been shown to perform slightly better

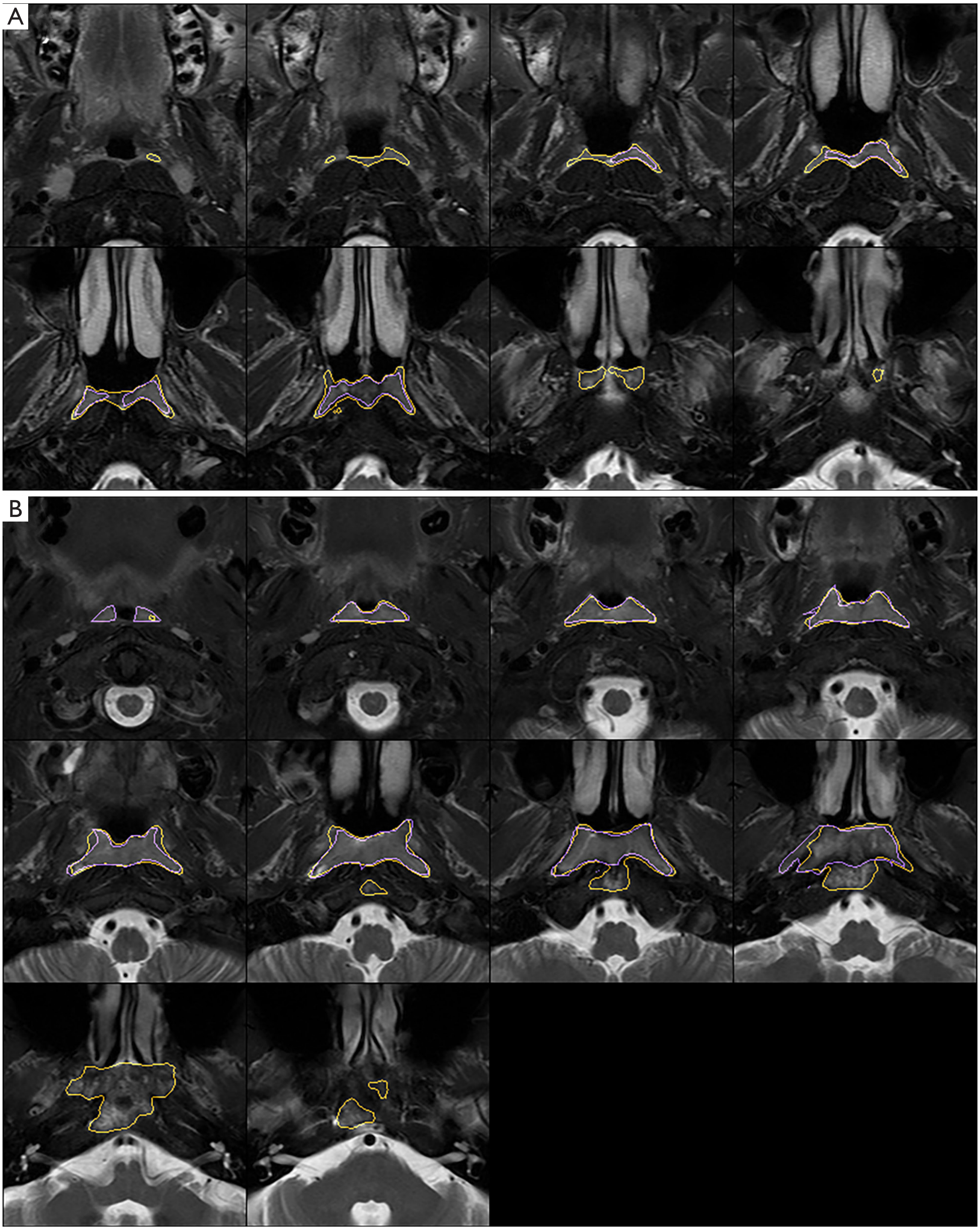

Figure 4 Primary tumour delineation overlaying the T2W-FS images of two cases which showed disagreement between CNN delineation $A_{proposed}$ (yellow) and first manual delineation M_{1st} (purple). (A) For early-stage primary tumours, the CNN tends to over-contour the primary NPCs where the tumours become thinner on the slices most distal to the centre of the tumour. (B) For advance stage primary NPCs, the ballooning of the sphenoid sinuses back to the clivus caused susceptibility artefacts in the air-bone interfaces that were mistakenly labelled as tumour by the CNN. T2W-FS, T2-weighted fat-suppressed; CNN, convolutional neural network; NPC, nasopharyngeal carcinomas.

Table 4 Median of performance metrics grouped by T-stage using the first set of manual delineation M_{1st} as the referenced ground-truth

Metrics	T1 (n=130)	T2 (n=59)	T3 (n=140)	T4 (n=75)	P value
$A_{proposed}$					
DSC	0.79 (0.10)	0.79 (0.11)	0.80 (0.09)	0.80 (0.06)	0.679
CR	0.71 (0.18)	0.69 (0.16)	0.69 (0.15)	0.69 (0.10)	0.668
PM	0.91 (0.15)	0.86 (0.15)	0.81 (0.15)	0.76 (0.13)	<0.001*
ASD (mm)	0.48 (0.75)	0.58 (0.79)	0.70 (0.74)	0.97 (1.28)	<0.001*
M_{2nd}					
DSC	0.82 (0.07)	0.81 (0.07)	0.81 (0.06)	0.82 (0.05)	0.386
CR	0.72 (0.11)	0.71 (0.11)	0.71 (0.10)	0.72 (0.09)	0.763
PM	0.80 (0.11)	0.79 (0.13)	0.83 (0.11)	0.80 (0.10)	0.028*
ASD (mm)	0.31 (0.42)	0.59 (0.83)	0.73 (0.74)	0.73 (0.97)	<0.001*

Data displayed are median (IQR). *, marks significant difference with the proposed method (P<0.05). DSC, Dice similarity coefficient; CR, correspondence Ratio; PM, percentage match; ASD, average surface distance; M_{1st}, manually drawn delineations set used as reference standard; M_{2nd}, 2nd manually drawn delineations set for intra-observer variability measurements; $A_{proposed}$, delineations generated by the proposed automatic CNN algorithm.

than 2D Dense-Unet (DSC =0.80 and 0.82 respectively), but we were unable to test H-Dense-Unet because it requires 3D isometric input while we only have anisometric 2D patches input. Interestingly all the CNNs as well as the human expert, encountered greater problems with specificity than with sensitivity, as reflected in the lower values for CR than PM. This suggests that the CNN, in common with human performance, was able to detect lesions with high sensitivity but had greater difficulty discriminating the aetiology of a detected lesion (i.e., benign or malignant), resulting in lower specificity.

Comparison of the proposed CNN with other CNN studies in the literature

We applied our proposed CNN to T2-weighted non-contrast-enhanced images with screening in mind. Only two previous studies have evaluated CNN primary NPC delineation using non-contrast-enhanced MRI (11,12). One study, using their CNN designed based on the dense-block technique, reported a DSC of 0.72 (12) and another, using the Unet, reported a maximum DSC of 0.65 (11), both of which showed lower performance than our $A_{proposed}$ (DSC =0.79). All other CNN studies for primary NPC delineation have reported the results in contrast-enhanced MRI using CNNs customised from Unet or Dense-Unet for delineating primary NPC (9-14). When compared to using contrast-enhanced images, our non-contrast-enhanced method archived better or comparable performance to that reported in four studies (mean/median DSC of 0.72–0.79) (9-12), but was worse than that reported in two small studies of 30 patients (DSC of 0.85) (13) and 29 patients (DSC of 0.89) (14).

Limitations of this study

This study has some limitations. Firstly, the proposed CNN algorithm requires images centred on the nasopharynx, so the technique may not be applicable if the area of coverage is expanded to include nodal disease in the neck. Nevertheless, the proposed CNN algorithm would offer improvements as long as the FOV coverage remains consistent. This aligns with clinical practice because each type of cancers usually has its routine MRI protocol which includes standard positioning of the FOV. Secondly, CNNs tend to smooth the boundaries of very irregularly shaped tumours, which can reduce the accurate contouring of the tumour boundary. Thirdly, as we performed the test on uniform images from one centre only, the effect of alternative scan settings especially on textural analysis is currently unknown. Fourthly, as CNN training is very time consuming, we only evaluated the previously published CNNs on data from patients in fold 1. However, as there were no significant differences between the four folds of our proposed CNN, we believe that it is likely that the superiority of our CNN in fold 1 is representative of the expected results from the other folds. Fifthly, this study did not assess the variations of the proposed algorithm in primary NPC delineation by CNN on contrast-enhanced MRI. Nonetheless, our previous work showed that the well-established Unet displayed similar primary NPC delineation performance on NE-T2W-FS when compared to contrast-enhanced T1-weighted images, and only slightly worse performance when compared to contrast-enhanced T1-weighted fat-suppressed images (20). Lastly, in this study we were unable to address the clinical importance of the differences between the manual and automatic delineations because of the complex invasion patterns of NPC and substantial differences in radiosensitivity of the different surroundings normal tissues.

Conclusions

We have developed and presented a fully automatic CNN algorithm that achieved a median DSC of 0.79 and ASD of 0.66 mm for delineating primary NPCs on a non-contrast-enhanced MRI sequence. The results suggest that our proposed CNN algorithm can automatically delineate primary NPCs with a DSC close to the previously established standard on a non-contrast-enhanced MRI sequence. The performance of our CNN on a T2-weighted sequence has great potential for MRI screening programs and intra-treatment assessment.

Acknowledgments

This study was presented, in part, by the first author at the 19th International Cancer Imaging Society annual meeting, October 7-9, 2019; Verona, Italy.
Funding: None.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/qims-21-196). Dr. LMW reports this study was presented, in part, by the first author at the 19th International Cancer Imaging Society annual meeting, October 7-9 2019; Verona, Italy. Attending cost was covered partially by the Department of Imaging and Interventional Radiology of The Chinese University of Hong Kong. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (Approval ID: CIE-2019.709), requirements of written consents were waived owing to its retrospective nature.

References

1. Fourcade A, Khonsari RH. Deep learning in medical image analysis: a third eye for doctors. J Stomatol Oral Maxillofac Surg 2019;120:279-88.
2. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts HJWL. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin 2019;69:127-57.
3. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham: Springer, 2015:234-41.
4. Hu Y, Zheng Y. A GLCM embedded CNN strategy for computer-aided diagnosis in intracerebral hemorrhage. arXiv:1906.02040 [Preprint]. 2019 [cited 2020 Jan 20]. Available online: http://arxiv.org/abs/1906.02040
5. Tan J, Gao Y, Cao W, Pomeroy M, Zhang S, Huo Y, Li L, Liang Z. GLCM-CNN: gray level co-occurrence matrix based CNN model for polyp diagnosis. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2019:1-4.
6. Liu R, Lehman J, Molino P, Such FP, Frank E, Sergeev

A, Yosinski J. An intriguing failing of convolutional neural networks and the CoordConv solution. arXiv:1807.03247 [Preprint]. 2018. Available online: https://arxiv.org/abs/1807.03247

7. Oktay O, Schlemper J, Folgoc L Le, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention U-Net: learning where to look for the pancreas. arXiv:1804.03999 [Preprint]. 2018 [cited 2020 Jan 19]. Available online: http://arxiv.org/abs/1804.03999
8. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. arXiv:1709.07330 [Preprint]. 2017 [cited 2020 Jan 19]. Available online: http://arxiv.org/abs/1709.07330
9. Lin L, Dou Q, Jin YM, Zhou GQ, Tang YQ, Chen WL, Su BA, Liu F, Tao CJ, Jiang N, Li JY, Tang LL, Xie CM, Huang SM, Ma J, Heng PA, Wee JTS, Chua MLK, Chen H, Sun Y. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. Radiology 2019;291:677-86.
10. Ke L, Deng Y, Xia W, Qiang M, Chen X, Liu K, Jing B, He C, Xie C, Guo X, Lv X, Li C. Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. Oral Oncol 2020;110:104862.
11. Chen H, Qi Y, Yin Y, Li T, Liu X, Li X, Gong G, Wang L. MMFNet: A multi-modality MRI fusion network for segmentation of nasopharyngeal carcinoma. Neurocomputing 2020;394:27-40.
12. Ye Y, Cai Z, Huang B, He Y, Zeng P, Zou G, Deng W, Chen H, Huang B. Fully-automated segmentation of nasopharyngeal carcinoma on dual-sequence MRI using convolutional neural networks. Front Oncol 2020;10:166.
13. Ma Z, Wu X, Song Q, Luo Y, Wang Y, Zhou J. Automated nasopharyngeal carcinoma segmentation in magnetic resonance images by combination of convolutional neural networks and graph cut. Exp Ther Med 2018;16:2511-21.
14. Li Q, Xu Y, Chen Z, Liu D, Feng ST, Law M, Ye Y, Huang B. Tumor segmentation in contrast-enhanced magnetic resonance imaging for nasopharyngeal carcinoma: deep learning with convolutional neural network. Biomed Res Int 2018;2018:9128527.
15. King AD, Vlantis AC, Bhatia KSS, Zee BCY, Woo JKS, Tse GMK, Chan ATC, Ahuja AT. Primary nasopharyngeal carcinoma: diagnostic accuracy of MR imaging versus that of endoscopy and endoscopic biopsy. Radiology 2011;258:531-7.
16. King AD, Woo JKS, Ai QY, Chan JSM, Lam WKJ, Tse IOL, Bhatia KS, Zee BCY, Hui EP, Ma BBY, Chiu RWK, van Hasselt AC, Chan ATC, Lo YMD, Chan KCA. Complementary roles of MRI and endoscopic examination in the early detection of nasopharyngeal carcinoma. Ann Oncol 2019;30:977-82.
17. King AD, Vlantis AC, Yuen TWC, Law BKH, Bhatia KS, Zee BCY, Woo JKS, Chan ATC, Chan KCA, Ahuja AT. Detection of nasopharyngeal carcinoma by MR imaging: diagnostic accuracy of MRI compared with endoscopy and endoscopic biopsy based on long-term follow-up. Am J Neuroradiol 2015;36:2380-5.
18. King AD, Woo JKS, Ai Q-Y, Mo FKF, So TY, Lam WKJ, Tse IOL, Vlantis AC, Yip KWN, Hui EP, Ma BBY, Chiu RWK, Chan ATC, Lo YMD, Chan KCA. Early detection of cancer: evaluation of MR imaging grading systems in patients with suspected nasopharyngeal carcinoma. Am J Neuroradiol 2020;41:515-21.
19. Chan KCA, Woo JKS, King A, Zee BCY, Lam WKJ, Chan SL, Chu SWI, Mak C, Tse IOL, Leung SYMS-FSYM, Chan G, Hui EP, Ma BBY, Chiu RWK, Leung SYMS-FSYM, van Hasselt AC, Chan ATC, Lo YMD. Analysis of plasma Epstein-Barr virus DNA to screen for nasopharyngeal cancer. N Engl J Med 2017;377:513-22.
20. Wong LM, Ai QYH, Mo FKF, Poon DMC, King AD. Convolutional neural network in nasopharyngeal carcinoma: how good is automatic delineation for primary tumor on a non-contrast-enhanced fat-suppressed T2-weighted MRI? Jpn J Radiol 2021. [Epub ahead of print]. doi: 10.1007/s11604-021-01092-x.
21. Leyba K, Wagner B. Gadolinium-based contrast agents: why nephrologists need to be concerned. Curr Opin Nephrol Hypertens 2019;28:154-62.
22. Choi JW, Moon WJ. Gadolinium deposition in the brain: current updates. Korean J Radiol 2019;20:134-47.
23. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, Gershenwald JE, Compton CC, Hess KR, Sullivan DC, Jessup JM, Brierley JD, Gaspar LE, Schilsky RL, Balch CM, Winchester DP, Asare EA, Madera M, Gress DM, Meyer LR. editors. AJCC cancer staging manual. 8th ed. Cham: Springer International Publishing, 2017.
24. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 2006;31:1116-28.
25. Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on

Kullback discrimination of distributions. In: Proceedings of 12th international conference on pattern recognition. IEEE, 1994;1:582-5.

26. Verma M, Raman B. Local neighborhood difference pattern: a new feature descriptor for natural and texture image retrieval. Multimed Tools Appl 2018;77:11843-66.
27. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. arXiv:1912.01703 [Preprint]. 2019. Available online: https://arxiv.org/abs/1912.01703
28. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation. arXiv:1704.06857 [Preprint]. 2017 [cited 2020 Jan 23]. Available online: http://arxiv.org/abs/1704.06857
29. Mattiucci GC, Boldrini L, Chiloiro G, D'Agostino GR, Chiesa S, De Rose F, Azario L, Pasini D, Gambacorta MA, Balducci M, Valentini V. Automatic delineation for replanning in nasopharynx radiotherapy: what is the agreement among experts to be considered as benchmark? Acta Oncol 2013;52:1417-22.

Cite this article as: Wong LM, Ai QYH, Poon DMC, Tong M, Ma BBY, Hui EP, Shi L, King AD. A convolutional neural network combined with positional and textural attention for the fully automatic delineation of primary nasopharyngeal carcinoma on non-contrast-enhanced MRI. Quant Imaging Med Surg 2021;11(9):3932-3944. doi: 10.21037/qims-21-196

Appendix 1

Algorithm details

Discriminative patch sampling

To increase the probability that the sampled patches contain tumour and reduce the probability that the sampled patches contain air or out-of-view space, we used an intensity-based patch sampling strategy for both inference and training of the network. This strategy was based on the mean intensity of tumour on T2W-FS images being higher than the mean of whole image and it used weighted sampling based on prior knowledge as shown in the following equation.

Without differentiating whether the slice contained tumour or not, for each 2D slice $S_z \in \mathbb{R}^{W\times H}$ at the axial position z, N square patches $\{L_n \in \mathbb{R}^{D\times D}; n \in [0,N) \cap \mathbb{Z}\}$ were sampled by selecting N points $\{w,h\}_n$ within the box $R_z \subset S_z$ bounded by $\{w \in [D/2, W-D/2), h \in [D/2, H-D/2)\}$. The probability of the point being selected is proportional to its pixel intensity:

$$P(w,h;z) = \frac{R_z(w,h) - \min[R_z]}{\sum_{w,h}\left(R_z(w,h) - \min[R_z]\right)} \qquad [5]$$

We used the empirical tests to determine the optimal patch size (W=128 and H=128) and the maximum number of patches (N=75) for each of the slices during inference to ensure all the tissue-representing pixels were covered during network inference. Figure S1 is an example to illustrate the effect of the intensity-based sampling.

Reflection padding

Stacking multiple convolutional blocks causes shrinkage of the receptive field (3). This can be partially compensated by boundary padding prior to each convolutional operation. There are four mainstream padding techniques, namely zero-padding, boundary-padding, reflection-padding and warped-padding. The zero-padding is the one which is most broadly applied because its implementation is simple and effective. However, zero-padding is less compatible with patch-based algorithms as it leads to sharpening of the local contrast at the edges of a patch which induces incoherence in the stitched output. We therefore adopted reflection-padding in all convolutional layers, a simple and elegant method to mitigate the receptive field shrinkage and the local contrast sharpening (Figure S2).

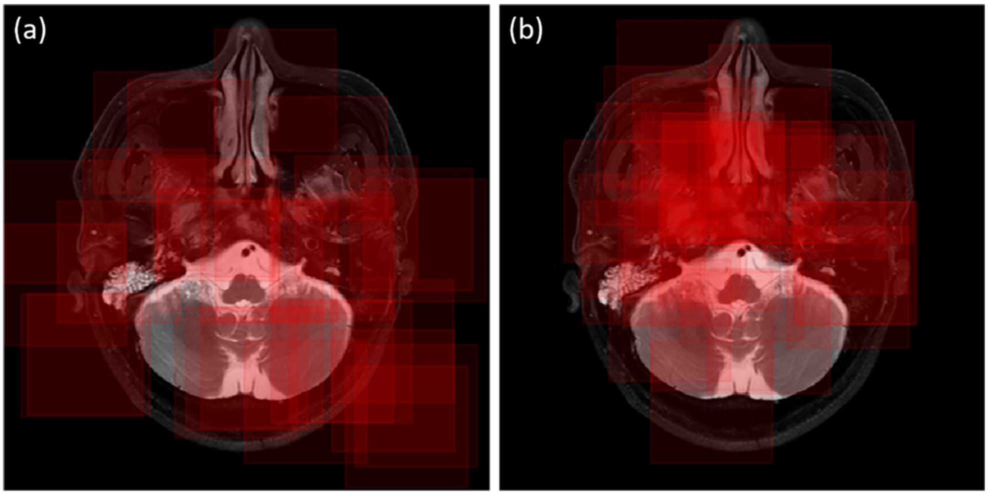

Figure S1 An example showing random sampling (A) and intensity-based sampling (B) on T2W-FS images. By using the intensity-based sampling method, sampled patches were more likely to focus on the nasopharynx and adjacent areas, whereas by using the random sampling method, sampled patches were scattered and included out-of-view space. T2W-FS, T2-weighted fat-suppressed.

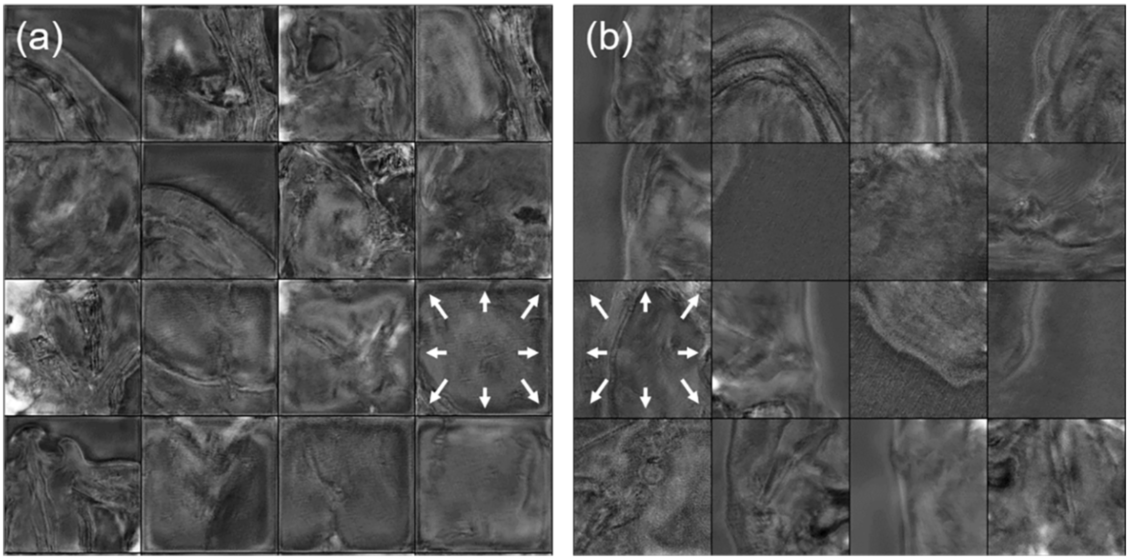

Figure S2 Compared to using convolution layers (A) with zero padding which shows severe distortion at the edge (white arrows), the use of (B) reflection-padding produced a smoother probability map without distortion at the edges (white arrows).

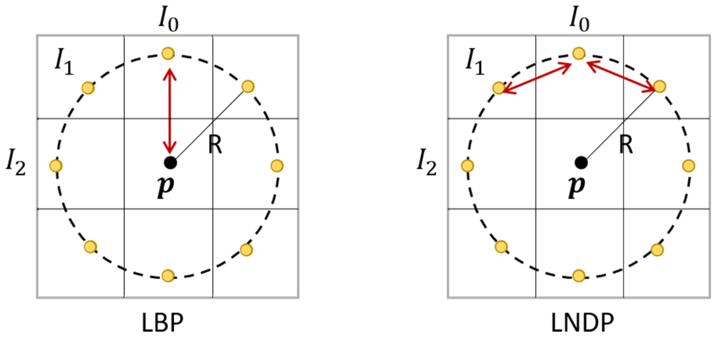

Figure S3 A pictorial explanation of the complementary relation between the LBP and the LNDP texture features. The LBP describes the local radial differences whereas the LNDP describes the local tangential differences. The frequencies of these differences were encoded into the centre pixels of each texture kernel in the output of these textural filters. R is the kernel radius, I_i is the i-th intensity in the intensity vector $I(K_p)=\{I_0, I_1…I_N\}$. The red double-headed arrows indicate the compared pair of points in each texture filter. LNDP, local neighbourhood differences pattern; LBP, local binary pattern.

Positional and textural feature vector

The routine NPC MRI scanning protocol was centred over the posterior wall of the nasopharynx. In order to locate the tumour, we recorded the 3D coordinates of each patch sampled and calculated the distance of each patch to the centre of the image. In addition, we introduced to the model two textural analysis techniques, namely the LBP (25) and LNDP (26), to improve the detection of tumour spread into adjacent soft tissues and bone, including those into areas less frequently affected where the 3D coordinates alone would have greater difficulty capturing the information. The frequencies of intensity differences were measured in a radial direction using the LBP and complementarily in the tangential direction using LNDP as shown in Figure S3.

A kernel radius of 3 px (~1.5 mm) was used for both textural analysis techniques. Two-hundred features were computed from the histograms of the LBP and LNDP texture of the extracted patches, and concatenated with the aforesaid 4-elements coordinate vector, forming a 204-elements feature vector that was the input of the fully-connected (FC) linear attention modules for each patch. Computation of LBP and LNDP from 2D images are recapped in Appendix 2.

Network architecture

The backbone of our network was based on the Attention-Unet (7), which adapted the Unet (3) with additional attention layers to capture semantic local information. We further replaced the convolutional attention layers with the FC linear attention layers which computes channel weights for each convolutional feature extractor layer. Specifically, it suppresses irrelevant feature extractor channels while reinforcing relevant ones according to the textural-positional vector of the extracted patch.

Appendix 2

Computation of LNDP and LBP

Details

For an input image, $I \in \mathbb{R}^{W \times H}$, each point $\boldsymbol{p}=\{w,h\}$ was iterated by considering M neighboring points $K_p=\{w+R\sin(2m\pi/M),h+R\cos(2m\pi/M);m\in[0,M)\cap\mathbb{Z}\}$ with fixed distance R from the considered point. It is common to select M to be an integer that is a power of two due to the bit-based data structure of digital images storage. A vector of intensities at points in the set K_p were obtained, denoted by $\boldsymbol{I}(K_p)\in\mathbb{R}^M$, where intensities at non-integer coordinates can be calculated with interpolation and those beyond the image boundaries can either be estimated with extrapolation or padding with little influence to the results.

For LBP, each element of the intensity vector $\boldsymbol{I}(K_p)$ was compared against that at the iterated point $I(p)$ with the following equations:

$$f(x,y)=\begin{cases}0 & \text{if} \quad x>y \\ 1 & \text{if} \quad \text{otherwise}\end{cases} \qquad [6]$$

$$LBP_I(\boldsymbol{p})=\sum_{m=0}^{M-1} f\left[\boldsymbol{I}(K_{\boldsymbol{p}})_m, I(\boldsymbol{p})\right]\times 2^m \qquad [7]$$

where $\boldsymbol{I}(K_p)_{\mathrm{m}}$ denotes the (m+1)-th element of the intensities vector.

For LNDP, difference between three consecutive elements in $\boldsymbol{I}(K_p)$ were compared.

$$g(x,y)=\begin{cases}0 & \text{if} \quad xy<0 \\ 1 & \text{if} \quad \text{otherwise}\end{cases} \qquad [8]$$

$$LNDP_I(\boldsymbol{p})=\sum_{m=0}^{M-1} g\left[\boldsymbol{I}(K_{\boldsymbol{p}})_{(m-1)\mathrm{mod}(M-1)}-\boldsymbol{I}(K_{\boldsymbol{p}})_m, \boldsymbol{I}(K_{\boldsymbol{p}})_{(m+1)\mathrm{mod}(M-1)}-\boldsymbol{I}(K_{\boldsymbol{p}})_m\right]\times 2^m \qquad [9]$$

We computed the LBP and LNDP images of the extracted patches with M=8 and R=1.5 mm (approximately 2 pixels from the centre or 3 pixels including the centre).

Selection of LBP and LNDP as additional features

In this study, LBP and LNDP were selected over other textural features, such as GLCM, because they are invariant to monotonic shifts of intensity values, which are common across clinical MRI scans. Furthermore, LBP and LNDP are a complementary textural filter pair which address local radial and local tangential intensity differences, respectively. Therefore, they provide both structural and statistical information when combined with histogram analysis.

Appendix 3

Pre-training data augmentation

Data augmentation referred to the perturbation of training data by applying transformations and/or signal filters to mimic the data variability encountered in real-life and improve the generality of the dataset. In this study, data augmentation was done prior to the patch extractions. Each 2D axial slice was augmented into two additional images using a pipeline described as follow:

(I) Random rotations between –10 to 10 degrees;
(II) Random scaling of the image to between 0.9 to 1.1 of its original size;
(III) Adding gaussian noise;
(IV) Linearly shifting the contrast of the images.

The upper augmenter pipeline was randomized after each epoch to maximize the robustness of the trained network. From each augmented image, 15 patches of dimension 128×128 px were sampled, the sampled locations were also randomized after each epoch.

Appendix 4

Network training

The open source python package PyTorch was used to implement and train the proposed network. The network was optimized using the Adaptive Moment Estimation (Adam) optimizer with the cross-entropy loss function on a machine with two graphic processing units (GPUs) model Nvidia 2080 Ti. The learning rate decayed exponentially after each epoch.

To account for smaller tumour compared to non-tumour volumes, we adopted class weightings in evaluating the loss function. The weights of tumour and non-tumour classes were calculated by the reciprocal of their total pixels counts and the non-tumour class weight was gradually increased using the sigmoid function across the epochs.

One epoch corresponded to a cycle in which the entire training cohort, including the augmented images, was passed forward through the network once. For a training cohort that included N axial slices, 45N patches would be sampled from 3N slices (2 augmented images and the source image) and processed by the network in each epoch.

http://dx.doi.org/10.21037/qims-21-196