Original Article

Context-aware augmentation for liver lesion segmentation: shape uniformity, expansion limit and fusion strategy

Qiang He[1], Yujie Duan[1], Zhiyu Yang[1], Yaxuan Wang[1], Liyu Yang[1], Lin Bai[2], Liang Zhao[1,2,3]

[1]Precision Medicine Research Center, Taihe Hospital, Hubei University of Medicine, Shiyan, China; [2]School of Computing and Electronic Information, Guangxi University, Nanning, China; [3]Hubei Key Laboratory of Embryonic Stem Cell Research, Hubei University of Medicine, Shiyan, China

Contributions: (I) Conception and design: L Zhao; (II) Administrative support: L Zhao , Q He; (III) Provision of study materials or patients: Y Duan, Z Yang, Y Wang, L Yang, L Bai; (IV) Collection and assembly of data: Q He; (V) Data analysis and interpretation: Q He; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Liang Zhao, PhD. Precision Medicine Research Center, Taihe Hospital, Hubei University of Medicine, 32 South Renmin Road, Shiyan, China. Email: s080011@e.ntu.edu.sg.

Background: Data augmentation with context has been an effective way to increase the robustness and generalizability of deep learning models. However, to our knowledge, shape uniformity, expansion limit, and fusion strategy of context have yet to be comprehensively studied, particularly in lesion segmentation of medical images.

Methods: To examine the impact of these factors, we take liver lesion segmentation based on the well-known deep learning architecture U-Net as an example and thoroughly vary the context shape, the expansion bandwidth as well as three representative fusion methods. In particular, the context shape includes rectangular, circular and polygonal, the expansion bandwidth is scaled by a maximum value of 2 compared to the lesion size, and the context fusion weighting strategy is composed of average, Gaussian and inverse Gaussian.

Results: Studies conducted on a newly constructed high-quality and large-volume dataset show that (I) uniform context improves lesion segmentation, (II) expanding the context with either 5 or 7 pixels yields the highest performance for liver lesion segmentation, depending on the lesion size, and (III) an unevenly distributed weighting strategy for context fusion is appreciated but in the opposite direction, depending on lesion size as well.

Conclusions: Our findings and newly constructed dataset are expected to be useful for liver lesion segmentation, especially for small lesions.

Keywords: Context-aware augmentation; shape uniformity; expansion limit; uneven fusion; liver lesion segmentation

Submitted Dec 19, 2022. Accepted for publication May 18, 2023. Published online Jul 05, 2023.
doi: 10.21037/qims-22-1399
View this article at: https://dx.doi.org/10.21037/qims-22-1399

Introduction

Liver lesion segmentation of medical images generated by, for example, ultrasonography, computed tomography (CT) and magnetic resonance (MR), is to distinguish the underlining pixels of lesions from normal tissues (1-3). Fishing out lesions is critical to disease diagnosis and therapy, particularly for those with tiny size, i.e., less than 1 square centimeter in practice (4,5). It has been widely affirmed, as well as confirmed, that operating on lesions at their very early stage (having a very small size) can yield

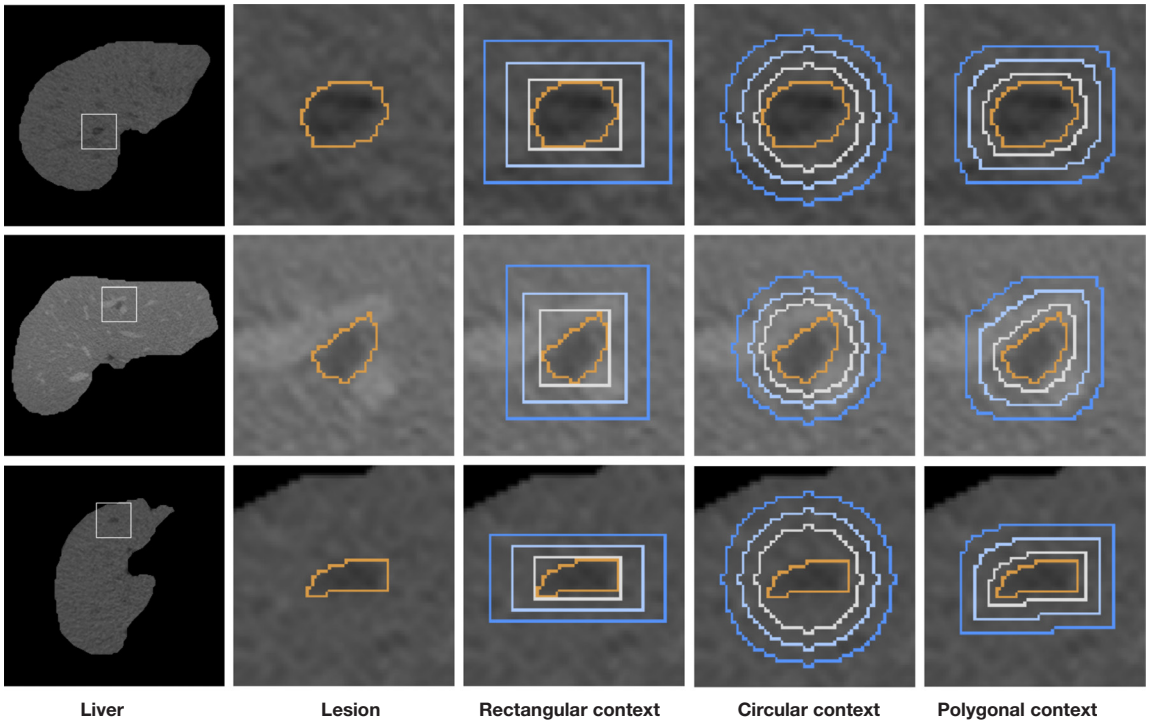

Figure 1 Illustration of various lesion contexts. The areas in orange are lesions, while the banded regions between the blue and orange lines are the context. The context expansions are confined by the blue lines with various intensities, and the uniformity is revealed by the shape.

positive results, even radically clearing them (6). However, segmentation of small lesions is very challenging due to the lack of adequate lesion areas for feature learning. Furthermore, there is a pressing need for sufficient samples for model training, especially tiny lesions.

To alleviate the constraints, various augmentation approaches have been proposed and applied to deep neural network training (7-10), which can be divided into two types, i.e., context-agnostic and context-aware augmentation. The context of an object is the surrounding environment in which the object is embedded. As shown in *Figure 1*, the blue and orange lines delineate the context in which the lesions are located. In context-agnostic augmentation, context is ignored; while in context-aware augmentation, the surrounding context is enclosed during augmentation.

Context-agnostic augmentation includes (I) simply copy and paste (8,11-13) and (II) mixing up objects with their labels (9,14-19). These methods have been proven to be effective in promoting the performance of semantic segmentation. However, the lack of context can be preventive to further improve these models' predictive ability (20,21). To address this, context-aware augmentations are proposed. The most widely used approach to achieve this goal is augmentation by a bounding box, i.e., repeating the object based on its bounding box or an enlarged bounding box (22,23). This type of augmentation is simple to implement, but the context is not equally distributed. Referring to *Figure 1* again, the context is significantly richer along the diagonal than parallel to the axes when the rectangular context is used.

Various context shapes and expansions are possible (*Figure 1*). These characteristics, to our knowledge, have yet to be studied. Inspired by the observation, we try to figure out the answers to these questions: (I) to what extent does the context contribute to semantic segmentation, (II) which context shape is most helpful, (III) how large the context should be expanded and, (IV) how to fuse a context into a background?

To answer these questions, we have conducted extensive experiments for liver lesion segmentation on a newly constructed high-quality and large-volume dataset. Our experiments are carried out strictly according to the logic of the questions and gradually moving from the initial scenario to the in-depth scenario. We have also performed experiments on the most widely used dataset LiTS (24).

Table 1 Details of LiTS and LSM

Dataset	#Volumes	#Slices	#Liver slices	#Lesions slices	#Lesions†	#Small lesions
LiTS	131	58,638	19,156	7,190	18,863	8,831
LSM	706	264,861	91,283	31,477	48,858	25,909

†, the number of lesions is determined at the slice level, not the volume level.

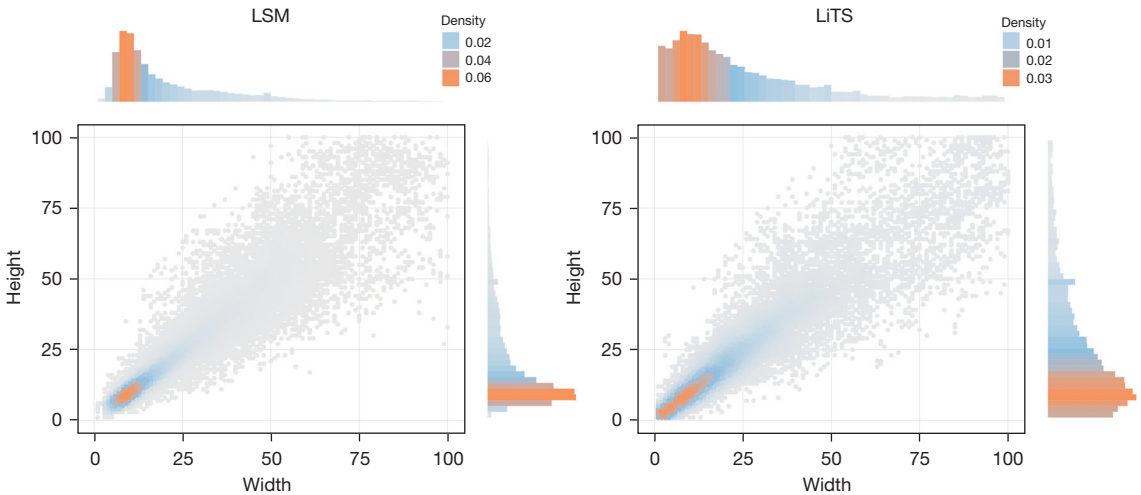

Figure 2 Detailed lesion size distribution of LiTS and LSM. The width and height of lesions are measured in pixel.

Our data, as well as the source codes, are available at https://github.com/lzhLab/LSM.

Methods

Datasets

Two datasets are used in this comprehensive study: the widely used dataset LiTS (24) and a newly constructed dataset LSM (short for liver lesion segmentation masks).

LiTS contains 131 volumes and 58,638 slices, of which 18,863 slices have lesions. Among these lesions, 8,831 have an area smaller than 1 square centimeter, which are deemed as small lesions, while the rest are large lesions. In terms of lesions at the volumetric level, there are 593 lesions. The classification of the lesion size is determined in accordance with (25), where small objects are defined as those whose size is less than 32*32 pixels.

LSM has 706 volumes and 91,283 liver-containing slices. In total, there are 48,858 lesions at the slice level (6,623 lesions at the volume level), of which 25,909 are small. To construct this high-quality and large-volume dataset, the lesions are delineated by three radiologists, and the final masks are the majority vote of the three. In case the consistency of a lesion mask is less than 0.5, the inconsistent one(s) will be sent back and refined again. The consistency is calculated as the ratio between the majority vote and the constituent mask. This dataset consists of various thicknesses. In particular, 17,671 slices are in 0.625 mm, 7,217 are in 1 mm, 66,068 are in 1.25 mm, 283 are in 2.5 mm, and 44 are in 5.0 mm.

The overview of the two datasets is shown in *Table 1*, while the detailed lesion size distributions are in *Figure 2*.

Please be noted that, although the classification of lesion size is in 2D, it is similar to the situation in 3D as most lesions in LiTS and LSM are very small and only occupy two to three slices; see *Figure 2*. Of course, 3D classification will give more accurate results.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). As this study does not involve human subjects, ethical approval and informed consent are not required.

Context determination

Three classical shapes of lesion-specific context are

explored, including rectangular, circular, and polygonal contexts. These three shapes represent increasingly better context uniformity but more burdensome implementation complexity.

The rectangular context is achieved by excluding the lesion from its bounding box and scaling it with a factor. The width and height of the bounding box equal the size of the lesion determined along the axes, while scaling is carried out by varying with the aspect ratio unchanged. Generally, is no less than 1, and a larger value indicates more context information.

For a circular context, the diameter is the maximum between the width and height of a lesion, and the center is the geometric mean of the bounding box. Similar to the rectangular one, the context can be enlarged by scaling the diameter by a factor α as well. Clearly, the context is more equally distributed along a lesion's boundary than the rectangular one (*Figure 1*).

The polygonal context is achieved by convolving a lesion with a kernel. Let $I \in R^{H\times W}$ be an image with size $H\times W$, I_{xy} be the intensity of the pixel at position (x,y) and I^e be a lesion contained in I, i.e., $\forall I^e_{xy} \in I$. The mask of I^e is $M \in R^{H\times W}$ with $M_{xy} \in \{0, 1\}$. M_{xy} equals 1 if (x,y) $\in I^e$, otherwise 0. To achieve the shape-specific context expansion on I^e, a circular kernel $k(x,y;r)$ filled with 1s is convoluted on M by

$$M'_{xy} = \max_{(i,j):K(i,j;r)=1} M\left(x+i, y+1\right) \quad [1]$$

where r is the radius of the kernel that is flexible during augmentation. Based on M', the expanded region $I^{e\prime}$ can be obtained by I′= I ⊙ M', where ⊙ is the matrix dot production.

Context-aware lesion embedding

Embedding a context-aware lesion into another place of a liver is to replace the target region with the lesion itself and fuse the context with the surrounding background.

Suppose a lesion I^e is to be embedded into a liver-contained image at position (x,y), and the original mask as well as the context-aware mask of the lesion are M and M', respectively. Then the lesion-embedded image I′ is

$$I'_{xy} = \begin{cases} I_{xy} & \text{if } M'_{xy} = 0 \\ \lambda_{xy} I^e_{xy} + \left(1-\lambda_{xy}\right) I_{xy} & \text{if } M'_{xy} - M_{xy} = 1 \\ I^e_{xy} & \text{if } M_{xy} = 1 \end{cases} \quad [2]$$

where λ_{xy} is the weight of the context at position (x,y) surrounding the lesion. Three weighting strategies are examined, including uniform, Gaussian and inverse Gaussian.

For the uniform weighting strategy, λ equals 0.5 at every position.

Regarding Gaussian, $\lambda = N\left(d\left(x,y\right);\mu,\sigma^2\right)$, d(x,y) is the Manhattan distance between the interested point (x, y) and the nearest point (x′, y′) of M having M(x′, y′) =1, and $N(\cdot)$ is the normal distribution. In this study, μ is set to 0 and σ is optimized as

$$\sigma = \left\lfloor \frac{d_0+1}{2} \right\rfloor / \sqrt{2\ln 2} \quad [3]$$

where d_0 is the maximal distance between the expanded pixels to the nearest lesion, i.e., $d_0 = \max \min d\left(x,y\right)$. For inverse Gaussian, λ is $1-N\left(d\left(x,y\right);\mu,\sigma^2\right)$ Analogously, μ is 0, and σ is the same as the Gaussian one.

Note that the above three approaches cover all the typical weighting strategies, in which the uniform one says the influence of context is equal at every position, the Gaussian one gives higher weight to the pixels that are closer to the lesion while lower those far away, and the inverse Gaussian is precisely the opposite of the Gaussian one. The context of the inverse Gaussian weighted lesions varies the most, while the Gaussian weighted ones the least.

Context-aware lesion augmentation

Two types of operations are applied to context-aware lesion augmentation to increase the number of lesions, including scale and rotation. The scale comprises scale-down and scale-up. Small lesions will be enlarged by scale-up, while large ones will be shrunk by scale-down. The two operations are carried out based on a predefined probability. Regarding rotation, any angle between 0 and 360 degrees can be randomly selected. Lesions and their surrounding context are rotated according to the randomly sampled angle correspondingly. Although there have many other image operators, such as crop, shift, and shear, they are ignored here because new images generated by these operations are impractical.

Results

Evaluation metrics

Five measurements, which are widely used in medical image

Table 2 Base-line performance of liver lesion segmentation on LiTS and LSM by using FCN, U-Net and DeepLabv3+

Data	Size	Model	DSC↑	VOE↓	ASSD↓	MSD↓	RMSE↓
LiTS	Small	FCN	0.444	0.346	7.643	2.684	0.042
		U-Net	0.468	0.320	6.525	2.344	0.038
		DeepLabv3+	0.412	0.340	8.185	2.826	0.043
	Large	FCN	0.756	0.301	15.54	4.868	0.112
		U-Net	0.768	0.284	14.46	4.644	0.106
		DeepLabv3+	0.767	0.288	14.51	4.678	0.109
LSM	Small	FCN	0.707	0.257	2.939	1.129	0.024
		U-Net	0.717	0.232	2.616	1.022	0.023
		DeepLabv3+	0.720	0.217	3.038	1.186	0.025
	Large	FCN	0.832	0.197	9.690	3.361	0.082
		U-Net	0.857	0.166	9.075	2.994	0.076
		DeepLabv3+	0.830	0.206	9.813	3.456	0.084

DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

segmentation evaluation (26), are borrowed, including dice similarity coefficient (DSC), volume overlap error (VOE), average symmetric surface distance (ASSD), maximum symmetric surface distance (MSD) and root mean square error (RMSE). They are defined as:

$$DSC(P,G)=\frac{2|P|\cap|G|}{|P|+|G|} \quad [4]$$

$$VOE(P,G)=1-\frac{|P\cap G|}{|P\cup G|} \quad [5]$$

$$ASSD(P,G)=\frac{1}{Z}\left(\sum_{x\in S(P)} d\left(x,S(G)\right)+\sum_{x\in S(G)} d\left(x,S(P)\right)\right) \quad [6]$$

$$MSD(P,G)=\max\left(\max_{x\in S(P)} d\left(x,S(G)\right),\max_{x\in S(G)} d\left(x,S(p)\right)\right) \quad [7]$$

$$RMSE(P,G)=\sqrt{\frac{1}{wh}\sum_{i=1}^{wh}\left(P_i-G_i\right)^2} \quad [8]$$

where P is the prediction, G is the ground truth, $Z = |S(P)|+|S(G)|$, $S(X)$ is the boundary of the region X, and $d(x, S)$ is the distance between any point x and a boundary S with $d(x, S) = \min_{y\in s} \|x - y\|$ and $\|\cdot\|$ the Euclidean distance.

The above metrics have covered all the main groups of image segmentation evaluation methods, i.e., spatial overlap, volume overlap, and spatial distance (24).

Baseline performance

To fairly compare the performance of various augmentation settings, baseline performance is obtained by applying three classical network architectures to the datasets with fivefold cross-validation. The three models are FCN (27), U-Net (28), and DeepLabv3+ (29), where the backbone is ResNet-34 (30). The detailed performance is shown in *Table 2*. Without specification, all the results shown here and beyond are evaluated at a volume level other than the slice level to eliminate validation bias.

Two main observations can be drawn from the results: (I) the performance obtained from LSM is markedly better than that from LiTS; (II) the performance generated from the large lesions is significantly better than the results yielded from the small ones. For the first observation, the superiority is mainly from the larger data volume, i.e., 707 *vs.* 131. In addition, the strict delineation protocol used in LSM construction can be beneficial to the improvements as well. Regarding the second, it consolidates the isolation of lesions by size during performance evaluation. Since the performance on LSM is markedly better than that on LiTS, we further conducted a cross-dataset analysis to examine the generalizability of the newly constructed dataset. That is, taking LSM for training and validation while using LiTS for testing, and vice versa. *Table 3* shows that the generalizability of LSM is significantly better than LSM.

Table 3 Generalizability analysis of LSM and LiTS

Train & Val.	Test	Size	Model	DSC↑	VOE↓	ASSD↓	MSD↓	RMSE↓
LiTS	LSM	Small	nnU-Net (2D)	0.443	0.258	1.473	3.327	0.025
		Large		0.477	0.266	4.751	12.58	0.025
LSM	LiTS	Small		0.653	0.497	7.839	3.041	0.029
		Large		0.767	0.358	26.41	7.386	0.083

DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error; FCN, fully convolutional network.

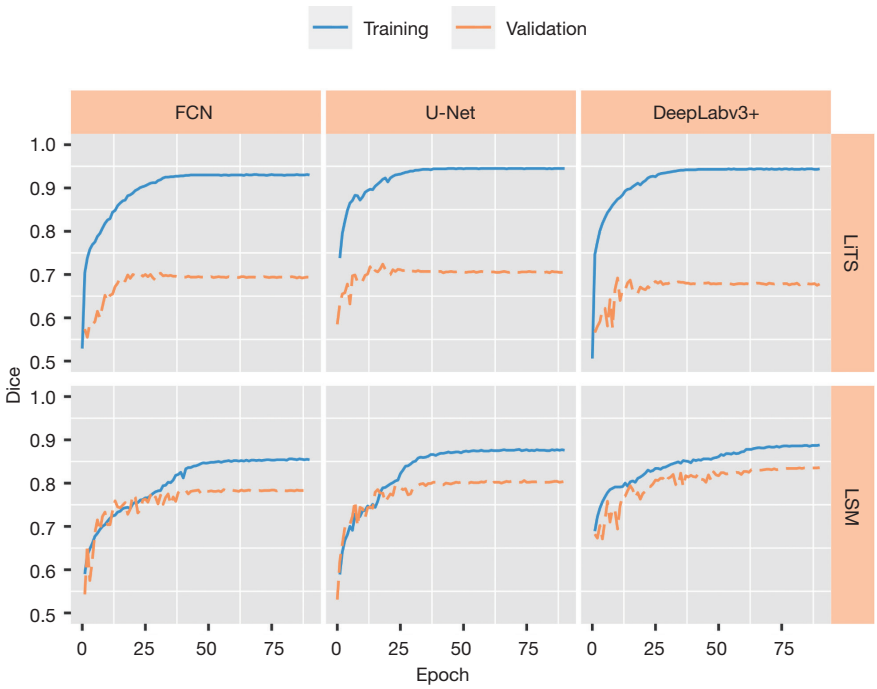

Figure 3 The training and validation dice of FCN, U-Net and DeepLabv3+ on LiTS and LSM. FCN, fully convolutional network.

In particular, the absolute Dice score is increased by 20% when LSM is used as training data compared to LiTS.

Figure 3 shows the detailed dice during the three models' training and validation on LiTS and LSM. After being well trained, the discrepancies between training and validation dice obtained from LiTS are significantly larger than that from LSM. This might be caused by the small amount of data with high heterogeneity contained in LiTS. This speculation is also supported by LiTS's faster convergence speed than LSM.

These observations indicate that preparing large-volume and high-quality datasets, like LSM, is necessary and helpful for liver lesion segmentation.

Herein only three classical and popular architectures of neural networks are employed because (I) they are representative and (II) this study focuses on context-aware augmentation analysis other than posting brand-new models. Among the three architectures, U-Net outperforms the other two in most cases. Hence, in the following analysis, U-Net is used to evaluate the performance under various scenarios on LSM without further specification.

Augmentation improves segmentation

We first ask whether augmentation is helpful for segmenting liver lesions. To this end, we performed copy and paste (CaP), scale, and rotation of the lesions contained in LSM, and carried out segmentation using U-Net. Results show

Table 4 Liver lesion segmentation performance under various augmentation methods.

Size	Method	DSC	VOE	ASSD	MSD	RMSE
Small	None	0.717	0.232	2.616	1.022	0.023
	CaP	0.770	0.230	2.458	0.970	0.023
	Scale	0.779	0.228	2.434	0.949	0.023
	Rotation	0.786	0.223	2.429	0.946	0.023
Large	None	0.857	0.166	9.075	2.994	0.076
	CaP	0.910	0.117	5.668	1.871	0.061
	Scale	0.915	0.115	5.476	1.800	0.061
	Rotation	0.916	0.119	5.766	1.907	0.060

Augmentation is only carried out once for each lesion with a predefined probability, and different augmentation methods are carried out separately. DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

Table 5 Segmentation performance under different copy numbers of lesions

Size	#Copy	DSC	VOE	ASSD	MSD	RMSE
Small	0	0.717	0.232	2.616	1.022	0.023
	1	0.770	0.230	2.458	0.970	0.023
	2	0.773	0.226	2.359	0.931	0.022
	3	0.766	0.233	2.461	0.978	0.023
	4	0.751	0.236	2.516	1.006	0.023
Large	0	0.857	0.166	9.075	2.994	0.076
	1	0.910	0.117	5.668	1.871	0.061
	2	0.922	0.114	5.581	1.835	0.061
	3	0.890	0.122	8.291	2.594	0.071
	4	0.887	0.132	8.850	2.790	0.074

DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

that augmentation increases segmentation performance by at least 5.6% in terms of dice score; see *Table 4*. In addition, scale and rotation are more effective than CaP in promoting segmentation accuracy, both for small and large lesions.

Next, we wonder how many times a lesion should be repeated. Thus, newly augmented datasets generated from LSM with lesions repeated from 0 to 4 times were constructed, and the same U-Net model was trained and tested. Results show that two-time replication yields better performance in terms of dice score (*Table 5*). This observation agrees with previous findings (25).

All the augmentation methods are randomly selected to increase lesion diversity at this stage. Hence, the full capability of different methods can be unveiled.

Context-aware augmentation further improves segmentation

Now we are interested in whether context information further improves the accuracy of lesion segmentation. To this end, we duplicate each lesion twice with its rectangular context enclosed and train the U-Net model with the same settings as the previous experiments. The results conducted on LSM show that context-aware augmentation yields

Table 6 Lesion segmentation performance with/without context

Shape	Size	DSC	VOE	ASSD	MSD	RMSE
None	Small	0.717	0.232	2.616	1.022	0.023
	Large	0.857	0.166	9.075	2.994	0.076
CaP	Small	0.773	0.226	2.359	0.931	0.022
	Large	0.922	0.114	5.581	1.835	0.061
Rec†	Small	0.791	0.222	2.356	0.929	0.023
	Large	0.930	0.111	5.485	1.843	0.061

†, rectangle enclosed context. Here the rectangle size is the same as the bounding box. CaP, copy and paste; DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

better performance, particularly for larger lesions. Precisely, the DSC score is lifted from 0.773 to 0.791 for small lesions (P value <2.2e−16), and this value is increased from 0.922 to 0.930 for large lesions (P value <2.2e−16). The detailed results are shown in *Table 6*.

Better uniformity yields higher accuracy

Context inclusion by bounding box is the most widely used strategy for context-aware augmentation because it is easy to implement. However, the dispersion of context obtained in this way is skewed because lesion shapes are rarely in rectangles, rather they are close to circles; cf. lesion shape distribution shown in *Figure 2*. To examine the effect of context uniformity on lesion segmentation, three types of context are investigated, i.e., rectangle-based (Rec), circle-based (Cir), and polygon-based (Ply). The rectangle-based context has the poorest uniformity, while the polygon-based context has the finest uniformity. In addition, to make a fair comparison, the context of the three shapes is determined so that they have a similar ratio between the pixels of context and lesions. Specifically, we calculate the average ratio between the number of pixels composing the rectangular context and the number of pixels within the lesions. This ratio is further used as a benchmark to determine the radius of circles as well as the bandwidth enclosing the lesions.

Experimental results show that various context inclusion methods generate different results, and the dice scores are consistently improved along with the refinement of context uniformity. In particular, polygon-based context generates the highest results, while rectangle-based context yields the poorest results. It is the same for both small and large lesions. See details in *Table 7*. The dices and losses obtained during training and validation shown in *Figure 4* also demonstrate the usefulness of context uniformity.

Context expansion has a limit

The context should not be too large or too small. When a context is expanded to include the entire liver, the augmentation collapses to mirror the original image; on the contrary, it is the reflection of the lesions themselves. Hence, we ask to what extent the context should be expanded.

To this end, we vary the context size and calculate segmentation performance. In particular, the rectangle-/circle-based context is expanded by a factor 1, 1.5 and 2, while the bandwidth of the polygon-based context is increased from 1 to 11 pixels in step 2. Results show that 1.5 times context expansion yields the highest performance compared to those with expansion factors 1 and 2. Regarding the polygonal context, 5- or 7-pixel bandwidth generates the best results for small and large lesions, respectively. In addition, either increasing or decreasing the context size will weaken the models' performance, as seen in *Table 8*. Detailed examples of lesion augmentation with the best expansion limit are shown in *Figure 5*.

Context fusion favors higher diversity

Intuitively, copy-and-paste a context-aware lesion from one place to another is irrational as the context may differ between the source and target, thus causing sudden changes. To mitigate this inconsistency, different weighting strategies for context fusion are examined, including uniform, Gaussian, and inverse Gaussian.

Table 7 Segmentation performance comparison with various context uniformity on LSM

Size	Shape	DSC	VOE	ASSD	MSD	RMSE
Small	CaP	0.770	0.230	2.458	0.970	0.023
	Rec	0.783	0.224	2.434	0.949	0.023
	Cir	0.786	0.218	2.347	0.925	0.023
	Ply	0.792	0.218	2.341	0.921	0.022
Large	CaP	0.910	0.117	5.668	1.871	0.061
	Rec	0.914	0.115	5.529	1.865	0.061
	Cir	0.917	0.115	5.516	1.837	0.061
	Ply	0.918	0.116	5.508	1.832	0.061

CaP, copy and paste; Rec, rectangular context; Cir, circular context; Ply, polygonal context; DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

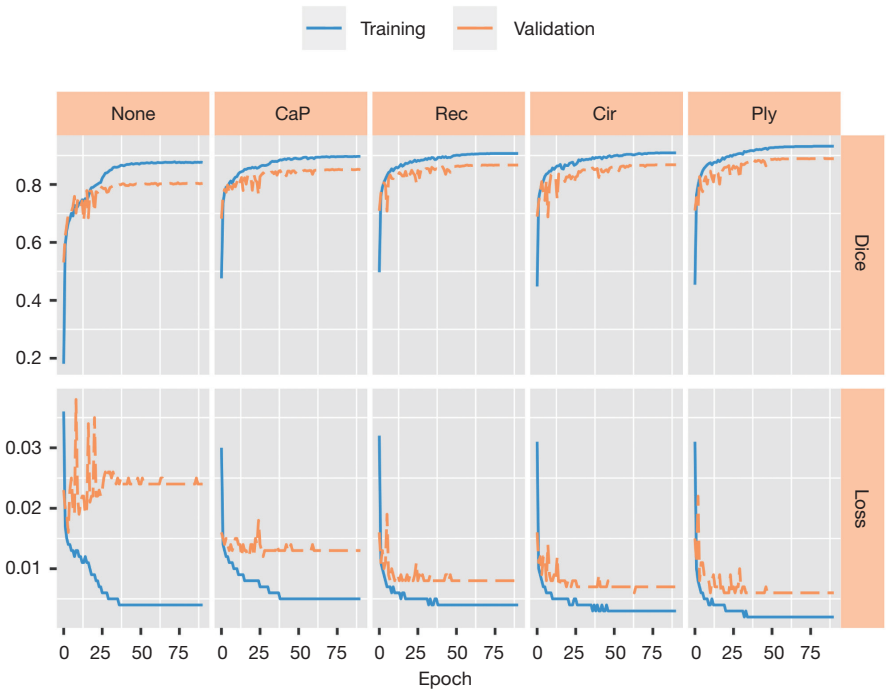

Figure 4 The training and validation dices and losses of lesion segmentation under various context uniformity. CaP, copy and paste; Rec, rectangular context; Cir, circular context; Ply, polygonal context.

Since the rectangular and circular context is not uniformly dispersed, and the performance obtained from these data is not as good as those generated from the polygonal context, they are ignored in this context fusion experiment. For polygonal context fusion, 5-pixel and 7-pixel bandwidths are applied to small and large lesions, as the best performance can be achieved under these conditions.

Results show that context fusion significantly improves the segmentation performance (*Table 9*). Interestingly, the inverse Gaussian-weighted context fusion generates the highest dice score for large lesions, while the Gaussian-weighted method produces the best results for small lesions. We speculate that small lesions are more sensitive to context. Hence, a lower variance produces better results.

Table 8 Lesion segmentation performance under various context expansions and uniformity

Context shape	Size	Extent	DSC	VOE	ASSD	MSD	RMSE
Rec	Small	1.0×[†]	0.783	0.224	2.434	0.949	0.023
		1.5×	0.788	0.220	2.374	0.936	0.023
		2.0×	0.772	0.229	2.540	0.977	0.024
	Large	1.0×	0.914	0.115	5.529	1.865	0.061
		1.5×	0.916	0.114	5.474	1.830	0.061
		2.0×	0.908	0.122	5.624	1.887	0.063
Cir	Small	1.0×	0.786	0.218	2.347	0.925	0.023
		1.5×	0.792	0.217	2.340	0.914	0.022
		2.0×	0.778	0.220	2.353	0.932	0.023
	Large	1.0×	0.917	0.115	5.516	1.837	0.061
		1.5×	0.918	0.114	5.461	1.817	0.061
		2.0×	0.913	0.116	5.668	1.871	0.062
Ply	Small	1p[‡]	0.792	0.218	2.341	0.921	0.022
		3p	0.796	0.218	2.336	0.920	0.023
		5p	0.803	0.216	2.318	0.911	0.022
		7p	0.797	0.218	2.340	0.920	0.023
		9p	0.796	0.218	2.346	0.924	0.023
		11p	0.790	0.219	2.421	0.928	0.024
	Large	1p	0.918	0.116	5.508	1.832	0.061
		3p	0.919	0.116	5.506	1.831	0.061
		5p	0.921	0.115	5.504	1.820	0.061
		7p	0.926	0.114	5.451	1.813	0.061
		9p	0.920	0.117	5.562	1.822	0.063
		11p	0.914	0.120	5.758	1.874	0.064

[†], context expansion scale factor; and [‡], bandwidth (in pixel) of context enclosing lesions. Rec, rectangular context; Cir, circular context; Ply, polygonal context; DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

To highlight the impact of different fusion strategies on context-aware lesion augmentation, we reveal the difference of context intensity in terms of log ratio between a fused pixel and the original one, see *Figure 6*. As can be seen, the Gaussian and reverse Gaussian fusion strategies apparently increase the diversity of lesion-surrounded context. Moreover, the unbalanced weighting strategy can preserve the original contextual continuity very well (white colored pixels indicate marginal changes in pixel intensity).

Performance with optimal components combined

On top of the gradual explorations, we come up with the final integration having context's shape, expansion and fusion considered. At the optimal parameter settings for different context shapes, the polygon-encircled context yields the best performance. In addition, context-aware models outperform context-agnostic models consistently. See *Figure 7*.

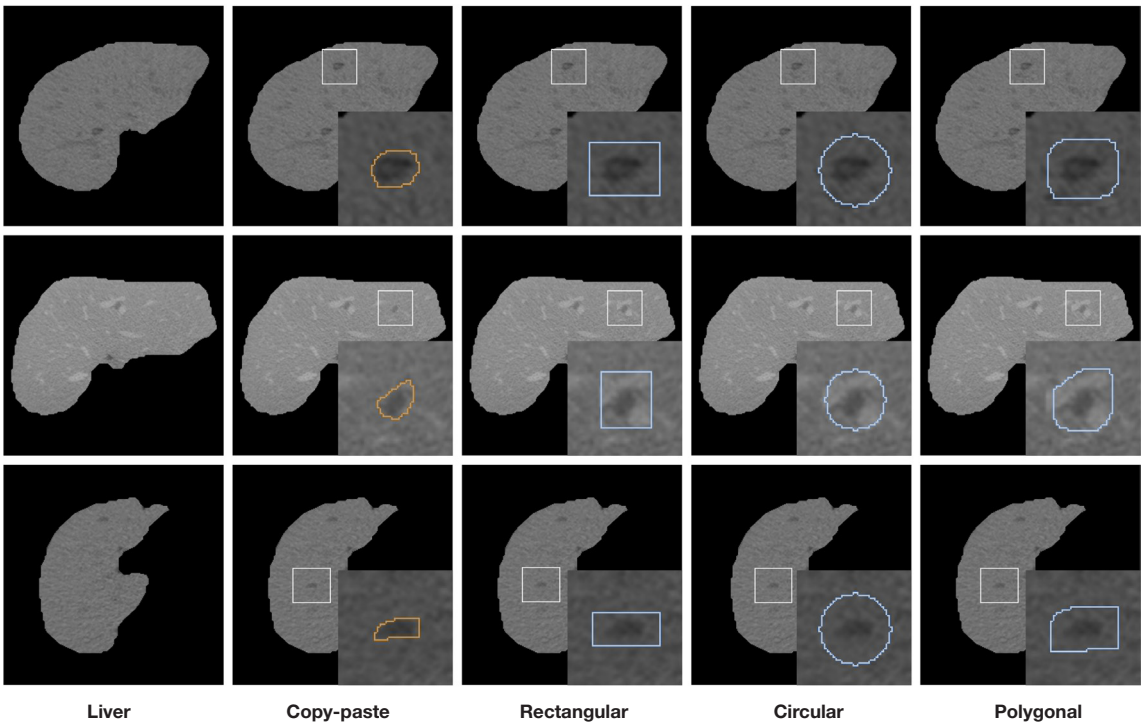

Figure 5 Liver lesion augmentation in various contexts. The augmented lesions are highlighted by a white box at the target region, while the details of lesions along with their surrounding context (in light blue) are zoomed in at the bottom-right corner.

Table 9 Segmentation performance comparison with Uniform, Gaussian and Rev Gaussian weighted context fusion strategies

Size	Weight	DSC	VOE	ASSD	MSD	RMSE
Small	Uniform	0.806	0.216	3.317	0.910	0.022
	Gaussian	0.824	0.214	3.228	0.902	0.020
	Rev Gaussian	0.813	0.215	2.310	0.907	0.021
Large	Uniform	0.930	0.111	5.297	1.754	0.061
	Gaussian	0.928	0.112	5.379	1.765	0.062
	Rev Gaussian	0.932	0.110	5.294	1.691	0.060

Rev Gaussian means reverse Gaussian weighting strategy. DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

Discussion

The context-aware augmentation analysis is mainly carried out on 2D images as the lesions are usually small, however, the strategy can be easily extended to 3D intuitively. To verify our hypothesis, we performed experiments on 3D volumes with and without context-aware augmentation by using nnU-Net (31) and MONAI (32) on LiTS and LSM, respectively. Results show that context-aware augmentation under the Gaussian fusion strategy improves segmentation performance remarkably; see *Table 10*. Moreover, the improvements in 3D are significantly larger than that in 2D. This is mainly due to the reduced lesion-to-liver ratio in 3D. Therefore, augmentation is particularly beneficial.

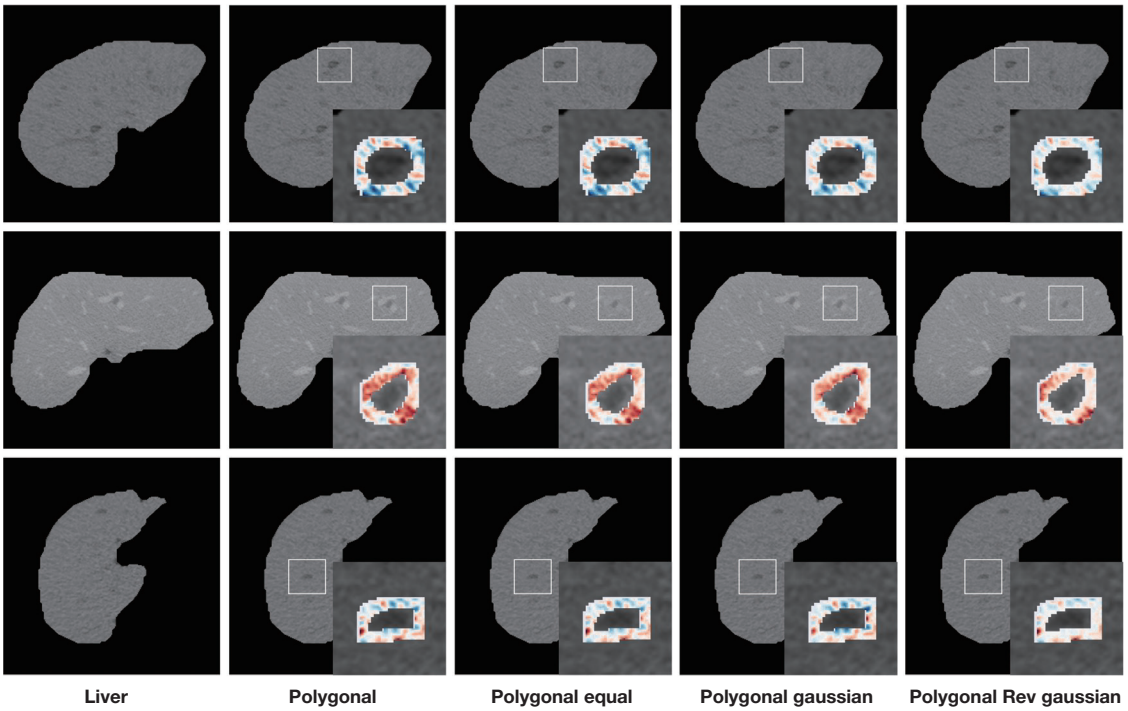

Figure 6 Impact of the context fusion strategy. The lesion-surrounded context regions are highlighted with color bands, where red indicates higher intensity than the original while blue represents the opposite. The color intensity is determined by the log ratio between a fused pixel and the original one.

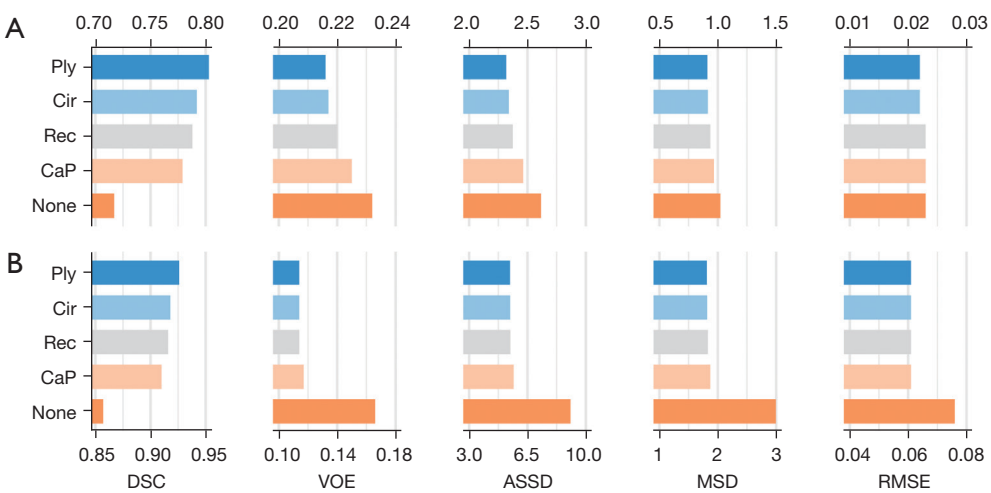

Figure 7 Performance on LSM under various context types combined with optimal settings. Panel (A) is for small lesions and panel (B) is for large lesions. DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

Table 10 Liver lesion segmentation performance comparison under various augmentation strategies achieved by 3D models

Data	Model	Aug	DSC↑	VOE↓	ASSD↓	MSD↓	RMSE↓
LiTS	nnU-Net	None	0.460	0.632	16.02	97.059	0.010
		Gaussian	0.775	0.338	2.571	39.270	0.038
	MONAI	None	0.385	0.292	77.27	271.29	0.032
		Gaussian	0.602	0.264	13.82	76.195	0.057
LSM	nnU-Net	None	0.767	0.350	5.787	63.297	0.008
		Gaussian	0.806	0.302	3.343	48.252	0.008
	MONAI	None	0.560	0.284	8.852	52.939	0.044
		Gaussian	0.728	0.201	4.840	55.397	0.039

DSC, dice similarity coefficient; VOE, volume overlap error; ASSD, average symmetric surface distance; MSD, maximum symmetric surface distance; RMSE, root mean square error.

Conclusions

Data augmentation with context has been proven helpful in semantic segmentation. Hence, it has been heavily used. However, existing context determination approaches mainly rely on an object's bounding box, which inevitably results in highly skewed context dispersion. To examine the effect of context, we comprehensively analyze the shape uniformity, expansion limit, and fusion strategy of context. We find that the polygonal context with the best context uniformity produces the highest accuracy in liver lesion segmentation, and the context should have a proper limit compared with its corresponding lesion size. In addition, an unevenly distributed weighting strategy for context fusion is more beneficial to lesion segmentation. Although the results above are drawn from liver lesion segmentation, the findings may shed light on other semantic segmentation tasks, particularly medical images.

Acknowledgments

Funding: This work was collectively supported by the National Natural Science Foundation of China (Nos. 32060150 and 61966003), the Advantages Discipline Group (Medicine) Project in Higher Education of Hubei Province (2021–2025) (No. 2022XKQT5), the Free Exploration Fund of Hubei University of Medicine (No. FDFR201805), and the Natural Science Foundation of Guangxi, China (No. 2020GXNSFAA159171).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1399/coif). The 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. The study does not involve any human subject, ethical approval and informed consent are not required. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Madalin M, Mircea S, Urhut CM, Sandulescu DL, Ionescu M, Streba CT. Liver lesion segmentation in contrast-enhanced ultrasound using deep learning algorithms. Ultrasound in Medicine and Biology 2022;48:S6.
2. Shi C, Xian M, Zhou X, Wang H, Cheng HD. Multi-slice low-rank tensor decomposition based multi-atlas segmentation: Application to automatic pathological liver

CT segmentation. Med Image Anal 2021;73:102152.
3. Zhao J, Li D, Xiao X, Accorsi F, Marshall H, Cossetto T, Kim D, McCarthy D, Dawson C, Knezevic S, Chen B, Li S. United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI. Med Image Anal 2021;73:102154.
4. Picon A, Terradillos E, Sánchez-Peralta LF, Mattana S, Cicchi R, Blover BJ, Arbide N, Velasco J, Etzezarraga MC, Pavone FS, Garrote E, Saratxaga CL. Novel Pixelwise Co-Registered Hematoxylin-Eosin and Multiphoton Microscopy Image Dataset for Human Colon Lesion Diagnosis. J Pathol Inform 2022;13:100012.
5. Liu Y, Li X, Li T, Li B, Wang Z, Gan J, Wei B. A deep semantic segmentation correction network for multi-model tiny lesion areas detection. BMC Med Inform Decis Mak 2021;21:89.
6. Nakamura S, Yamamoto T, Teng Y, Matsumoto S, Kasano K, Yoshiwara H, Hattori E, Tokunaga T, Yonetsu T, Hirao K. Impact of intensively lowered low-density lipoprotein cholesterol on deferred lesion prognosis. Catheter Cardiovasc Interv 2020;95:E100-7.
7. Pawar K, Egan GF, Chen Z. Domain knowledge augmentation of parallel MR image reconstruction using deep learning. Comput Med Imaging Graph 2021;92:101968.
8. Ghiasi G, Cui Y, Srinivas A, Qian R, Lin TY, Cubuk ED, Le QV, Zoph B. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2021:2917-27.
9. Hendrycks D, Mu N, Cubuk ED, Zoph B, Gilmer J, Lakshminarayanan B. Augmix: A Simple Data Processing Method to Improve Robustness and Uncertainty. In: International Conference on Learning Representations; 2020.
10. Yang M, Colak C, Chundru KK, Gaj S, Nanavati A, Jones MH, Winalski CS, Subhas N, Li X. Automated knee cartilage segmentation for heterogeneous clinical MRI using generative adversarial networks with transfer learning. Quant Imaging Med Surg 2022;12:2620-33.
11. Zhang JW, Zhang YC, Xu XW. ObjectAug: Object-level Data Augmentation for Semantic Image Segmentation. In: International Joint Conference on Neural Networks (IJCNN); 2021:1–8.
12. Fang HS, Sun JH, Wang RZ, Gou MH, Li YL, Lu CW. Instaboost: Boosting instance segmentation via probability map guided copy-pasting. In: IEEE/CVF International Conference on Computer Vision (ICCV); 2019:682-91.
13. Dwibedi D, Misra I, Hebert M. Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection. In: IEEE International Conference on Computer Vision (ICCV); 2017:1301-10.
14. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. AutoAugment: Learning Augmentation Policies from Data. ArXiv 2018; abs/1805.09501.
15. Gudovskiy D, Rigazio L, Ishizaka S, Kozuka K, Tsukizawa S. AutoDO: Robust AutoAugment for Biased Data with Label Noise via Scalable Probabilistic Implicit Differentiation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021:16601-10.
16. Kuo CW, Ma CY, Huang JB, Kira Z. FeatMatch: Feature-Based Augmentation for Semi-Supervised Learning. In: Vedaldi A, Bischof H, Brox T, Frahm JM, editors. Computer Vision – ECCV; 2020:479-95.
17. Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel CA. MixMatch: A Holistic Approach to Semi-Supervised Learning. In: Wallach H, Larochelle H, Beygelzimer A, d' Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32 (NeurIPS 2019); 2019.
18. Zhang HY, Cisse M, Dauphin YN, Lopez-Paz D. mixup: Beyond Empirical Risk Minimization. In: International Conference on Learning Representations; 2017.
19. Li B, Wu F, Lim SN, Belongie S, Weinberger KQ. On Feature Normalization and Data Augmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021:12383-92.
20. Hu PY, Ramanan D. Finding Tiny Faces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:1522-30.
21. Leng JX, Liu Y. Context Augmentation for Object Detection. Applied Intelligence 2021;52:1-13.
22. Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In: IEEE/CVF International Conference on Computer Vision (ICCV); 2019:6022-31.
23. Dvornik N, Mairal J, Schmid C. On the Importance of Visual Context for Data Augmentation in Scene Understanding. IEEE Trans Pattern Anal Mach Intell 2021;43:2014-28.
24. Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, et al. The Liver Tumor Segmentation Benchmark (LiTS). Medical Image Analysis 2023;102680.
25. Kisantal M, Wojna Z, Murawski J, Naruniec J, Cho K. Augmentation for small object detection. In: 9th International Conference on Advances in Computing and Information Technology 2019:119-33.

26. Nai YH, Teo BW, Tan NL, O'Doherty S, Stephenson MC, Thian YL, Chiong E, Reilhac A. Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. Comput Biol Med 2021;134:104497.
27. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:640-51.
28. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells AF William M and Frangi, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI; 2015:234-41.
29. Chen LC, Zhu YK, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. The European Conference on Computer Vision (ECCV) 2018:833-51.
30. He KM, Zhang XY, Ren SQ, Sun J. Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016:770-8.
31. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-11.
32. Diaz-Pinto A, Alle S, Ihsani A, Asad M, Nath V, Pérez-García F, Mehta P, Li WQ, Roth HR, Vercauteren T, Xu DG, Dogra P, Ourselin S, Feng A, Cardoso MJ. MONAI Label: A framework for AI-assisted Interactive Labeling of 3D Medical Images. Arxiv:2203.12362.

Cite this article as: He Q, Duan Y, Yang Z, Wang Y, Yang L, Bai L, Zhao L. Context-aware augmentation for liver lesion segmentation: shape uniformity, expansion limit and fusion strategy. Quant Imaging Med Surg 2023;13(8):5043-5057. doi: 10.21037/qims-22-1399

