Vector Quantisation for Robust Segmentation

Ainkaran Santhirasekaram[1], Avinash Kori[1], Mathias Winkler[2], Andrea Rockall[2] and Ben Glocker[1]

[1]Biomedical Image Analysis Group, Imperial College London | [2]Department of Surgery and Cancer, Imperial College London

Aim

Given an input x, we first define a function f(x) to represent the transformed input due to perturbation. The aim in this work is to find a way to learn a model (Φ) with weights w to be robust against δ(x) and construct an uncorrupted segmentation y from the perturbed input f(x).

Proposition

- The reliability of segmentation models in the medical do-main depends on the model's robustness to perturbations in the input space.
- Robustness is a particular challenge in medical imaging exhibiting various sources of image noise, corruptions, and domain shifts.
- We propose quantisation of the latent space of any segmentation network architecture, mapping the input images to a lower dimensional embedding space increases robustness to perturbation in the input space [1].
- We derive an empirically driven upper bound for maximum allowed shift in the latent space due to perturbation for robustness to hold.
- We focus on anatomical segmentation which benefits most from a quantized latent space.

Assumptions

- Assuming a small value for δ(x), we can then approximate Φ(x + δ(x)) with a first order Taylor expansion as follows: $\Phi(x + \delta(x)) = \Phi(x) + \delta(x)^T \nabla_x \Phi$. Therefore, the training framework should optimize for $\text{argmin}_w[\Phi(x+\delta(x)) - \Phi(x)]$ to be robust.
- In this work we assume that the segmentation network can be decomposed into an encoder (Φ_e) and decoder (Φ_d) such that $\Phi = \Phi_d \circ \Phi_e$, where $\Phi_e : X \rightarrow E$ maps from image space to a lower dimensional embedding space and $\Phi_d : E \rightarrow Y$ maps the embedding space back to segmentation space.

Quantisation for Robustness

- The quantisation process initially requires us to define a codebook $c \in R^{KxD}$. K is the size of the codebook and D is the dimensionality of each D codebook vector $l_i \in R^D$.
- We then define a discrete uniform prior and learn a categorical distribution P(z | x) as follows [1]:

$$\mathbb{P}(z = k \mid x) = \begin{cases} 1, & \text{for} \quad k = argmin_i ||\Phi_e(x) - l_i||_2 \\ 0, & \text{otherwise} \end{cases}$$

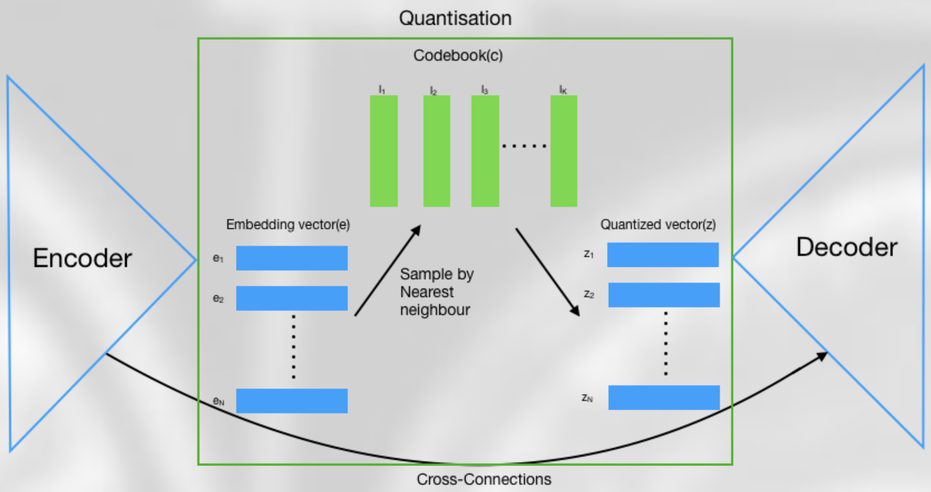

Figure 1: Proposed model architecture. We consider the UNet as our benchmark segmentation architecture and for the proposed architecture, VQ-UNet, we add a vector quantisation block at the bottleneck layer of the baseline UNet. Our codebook size (K) is 1024 each of dimension (D) 256. We consider both 2D and 3D Unets for 2D and 3D datasets respectively.

- Quantisation is non-differentiable, so we update the codebook weights with straight-through gradient approximation. We use the following loss function with stop gradient (sg) applied to constrain the update to the appropriate operand [1].

$$\mathcal{L}_{total} = \mathcal{L}_{Dice}(\hat{y}, y) + \mathcal{L}_{CE}(\hat{y}, y) + \|sg(\Phi_e(x)) - l\|_2 + \beta\|\Phi_e(x) - sg(l)\|_2$$

The first two terms in the above equation refers to the dice and cross entropy loss while the last two terms aims to reduce the Euclidean distance between the codebook vectors and the output of the encoder.

- Given the assumptions made, $\Phi_q(\Phi_e(x+\delta(x))) = \Phi_q(\Phi_e(x)+\delta(x)^T \nabla_x \Phi_e(x))$. We claim, quantisation pushes $\delta(x)^T \nabla_w \Phi_e(x)$ to 0 and thereby enforces $\Phi_q(\Phi_e(x + \delta(x))) = \Phi_q(\Phi_e(x))$.

Perturbation Bounds

- The maximum perturbation allowed around a single codebook vector denoted r and calculated empirically as half the average distance between a codebook vector (l_i) and its nearest neighbour (l_{i+1}) across the whole of the codebook provided the codebook is uniformly distributed. This is defined in the equation below as follows:

$$r = \frac{\sum_{i=0}^{i=K-1} \frac{1}{2}(\|l_i - l_{i+1}\|_2)}{K-1}$$

- We can the above equation and the first order Taylor expansion of $\Phi_e(x + \delta(x))$ to theoretically express r in terms of δ(x) as follows:

$$r > \|\delta(x)^T \nabla_x e_j\|_2$$

- Therefore, to affect an output of the quantisation block Φ_q, a perturbation δ(x) should lead to a change in the embedding space (e) greater than r whose upper bound is expressed in terms of δ(x).

Datasets

- We use the following 3 datasets for our experiments

	Abdomen	Prostate	Chest
Name	Beyond the Cranial Vault (BTCV) [2]	NCI-ISBI13 Challenge [3]	NIH Chest X-ray and the Japanese Society of Radiological Technology (JSRT) dataset [4].
Number of scans	30 CT scans	60 T2 weighted MRI scans of which half (BMC) are acquired on a 1.5T scanner with an endorectal coil and the other half (RUNMC) on a 3T scanner with a surface coil	100 Chest X-rays (NIH) 154 Chest X-rays (JRST)
Preprocessing	Normalised to 0-1 and resampled to 1.5×1.5×2mm	All images were re-sampled to 0.5×0.5×1.5mm and z-score normalized	Images were resized to 512×512 pixels and normalised to 0-1

Perturbation Study

- We compare how much the latent space changes in both models with different perturbations in the input space.
- We choose three different types of noise perturbations (Gaussian, salt and pepper, and Poisson noise) under 5 noise levels ranging from 0% to 30% to justify our claim of robustness.

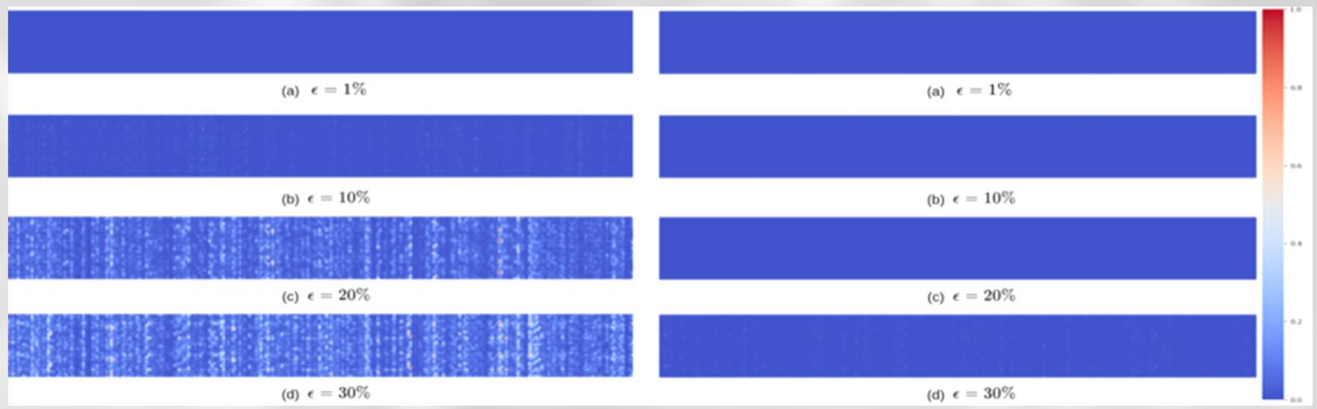

Figure 2: Variance heatmap of Unet (left) and VQ-Unet (right) latent space under 4 Gaussian noise levels for the NIH dataset. X-axis indicates a unique subset of features from a latent space, Y-axis corresponds to 100 randomly sampled test set images, and value at each location indicates the variance of a specific feature for a given image across 100 test time augmentations with Gaussian noise.

In Table 2, It can be seen that latent space features in VQ-UNet are not significantly changed (close to 0 variance) under various types of noise. The results are visualised in Figure 2 whereby the latent space of the VQ-UNet does not significantly change compared to the UNet under the addition of up to 30% Gaussian noise in the NIH dataset.

Table 2: Average latent space variance in both the models for all three datasets.

	Abdominal CT			Chest X-ray			Prostate		
	Gauss. Noise	S &P Noise	Poisson Noise	Gauss. Noise	S &P Noise	Poisson Noise	Gauss. Noise	S &P Noise	Poisson Noise
UNet	0.46	0.44	0.46	0.51	0.43	0.47	0.56	0.51	0.51
VQ-UNet	**3e-4**	**5e-5**	**2e-4**	**2e-4**	**1e-4**	**3e-4**	**1e-4**	**6e-5**	**8e-5**

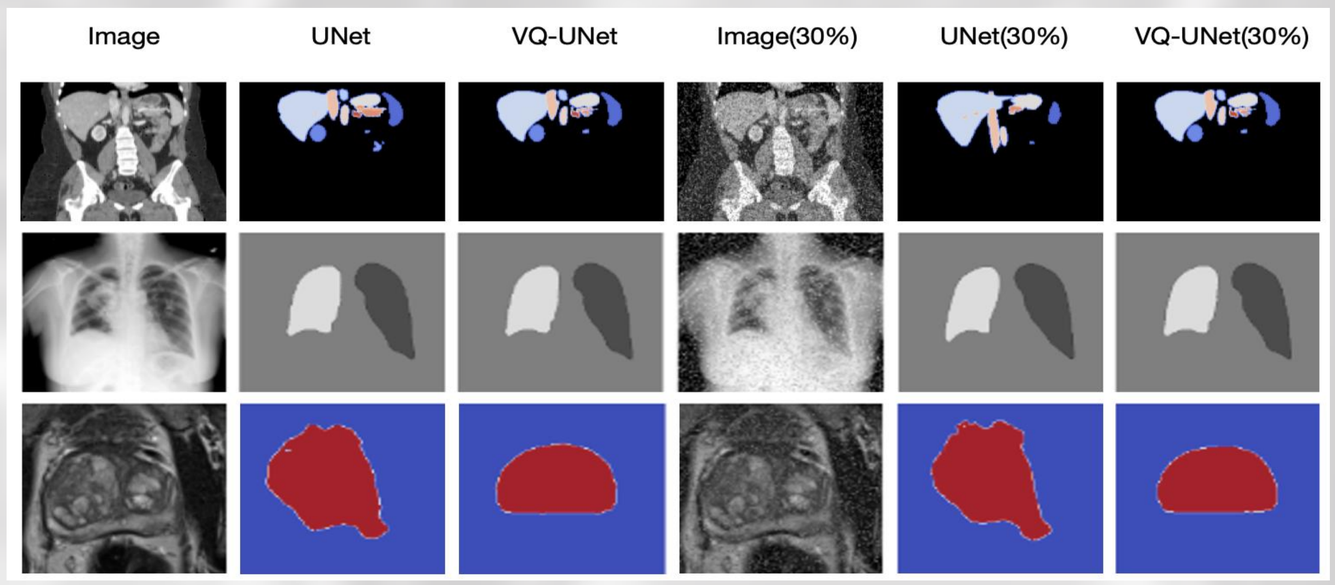

Figure3: Image and segmentation output for UNet and VQ-UNet under 0% (1st 3 columns) and 30% (2nd 3 columns) for s&p noise. Abdomen (1st row), Chest X-Ray NIH Sample (Second row), BMC Prostate Sample (Third Row).

We note in Figure 3 the segmentation maps produced by the VQ-UNet under the addition of 30% Gaussian noise do not change visually compared to the UNet. We make similar findings for salt & pepper noise and Poisson noise which is highlighted quantitively in Table 3.

Table 3: Gaussian noise perturbation on all 3 datasets

	Dice @0%	Dice @1%	Dice @10%	Dice @20%	Dice @30%
Chest X-ray NIH dataset					
UNet	0.95 ± 0.02	0.96 ± 0.02	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03
VQ-UNet	**0.97 ± 0.01**	**0.97 ± 0.01**	**0.97 ± 0.01**	**0.96 ± 0.02**	**0.96 ± 0.02**
Abdominal CT					
UNet	0.77 ± 0.01	0.76 ± 0.02	0.77 ± 0.04	0.76 ± 0.04	0.75 ± 0.08
VQ-UNet	**0.80 ± 0.01**	**0.79 ± 0.01**	**0.80 ± 0.01**	**0.80 ± 0.02**	**0.79 ± 0.02**
Prostate BMC dataset					
UNet	0.80 ± 0.02	0.81 ± 0.02	0.80 ± 0.03	0.78 ± 0.03	0.77 ± 0.06
VQ-UNet	**0.82 ± 0.02**	**0.82 ± 0.02**	**0.82 ± 0.02**	**0.82 ± 0.03**	**0.80 ± 0.04**

Domain Shift Study

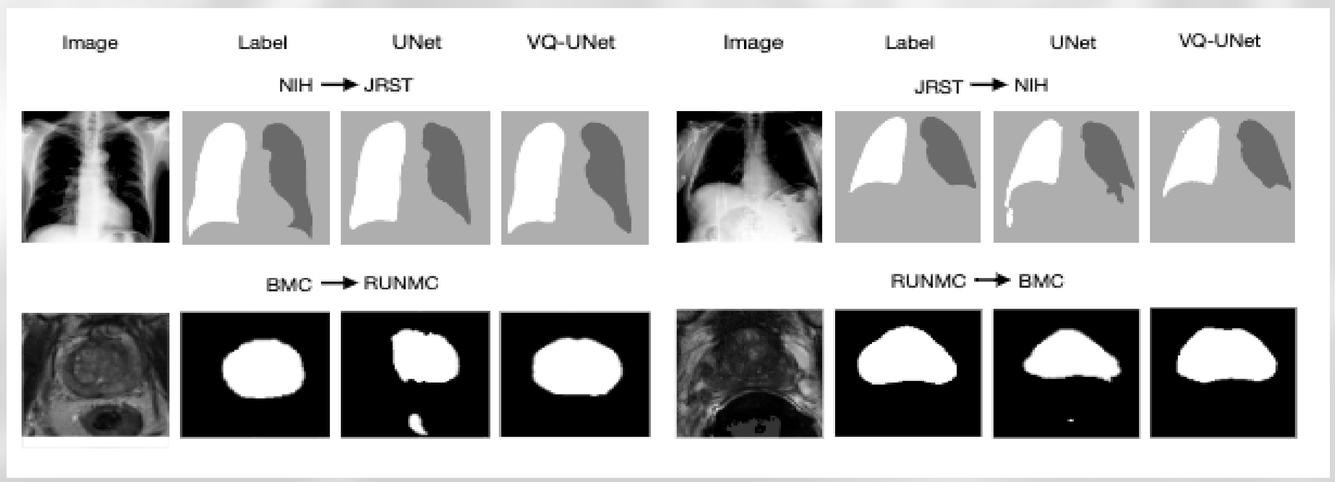

Figure 4: Sampled image input and Segmentation output for 2 domain shifts in chest X-ray (top row) and prostate (bottom row).

The VQ-UNet improved the segmentation performance both on the validation set and test set from a different domain for both prostate and chest X-ray. One notes particularly the ability to produce segmentation shapes which are more anatomically more meaningful. This highlighted quantitatively in the Dice score and and 95% Haussdorf distance in table 4.

Table 4: Mean Dice and HD95 on the validation sets for a single domain and test set across domain. The arrow represents the domain shift

	Chest X-ray							
	JRST		NIH		JRST→NIH		NIH→JRST	
	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95
UNet	0.93	7.31	0.96	6.80	0.95	7.12	0.82	8.27
VQ-UNet	**0.94**	**7.21**	**0.970**	**6.01**	**0.96**	**6.51**	**0.85**	**7.79**
	Prostate							
	BMC		RUNMC		BMC→RUNMC		RUNMC→BM	
	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95
UNet	0.80	8.42	**0.824**	7.84	0.55	33.3	0.62	25.7
VQ-UNet	**0.82**	**7.82**	0.822	**7.11**	**0.59**	**31.5**	**0.71**	**21.4**

Conclusion

- We propose and justify that given a segmentation architecture which maps the input space to a low dimensional embedding space, learning a discrete latent space via quantisation improves robustness of the segmentation model.
- In future work we propose to constrain the manifold upon which the embed the codebook. Specifically we hypothesize uniformly spreading the codebook vectors on the surface of the hypersphere improves robustness.

References

1. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. Advances in neural information processing systems 30 (2017)
2. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge. vol. 5,
3. Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K.: Nci-isbi 2013 challenge: automated segmentation of prostate structures. The Cancer Imaging Archive 370, 6 (2015).
4. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
5. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.i., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. American Journal of Roentgenology 174(1), 71–74 (2000)
6. Tang, Y.B., Tang, Y.X., Xiao, J., Summers, R.M.: Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation. In: International Conference on Medical Imaging with Deep Learning. pp. 457–467. PMLR (2019)

Acknowledgements:
This work was supported and funded by Cancer Research UK (CRUK) (C309/A28804) and UKRI centre for Doctoral Training in Safe and Trusted AI (EP/S023356/1).

Imperial College London

BioMedIA

MICCAI2022 Singapore