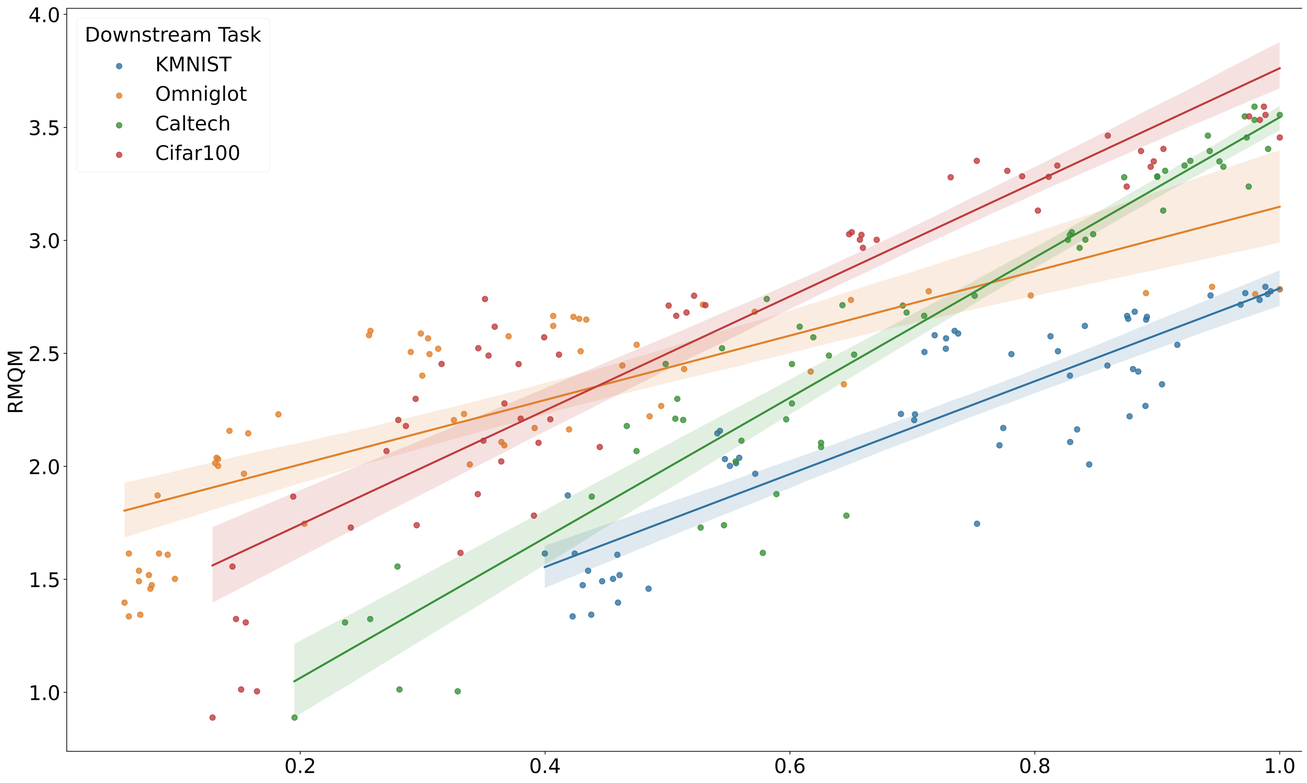
MANIFOLD CHARACTERISTICS THAT PREDICT DOWNSTREAM TASK PERFORMANCE

AUTHORS
Ruan van der Merwe | Greg Newman | Etienne Barnard

AFFILIATIONS
Bytefuse AI | Multilingual Speech Technologies, North-West University



Understanding why deep neural networks generalise so well remains a topic of intense research, despite the practical successes that have been achieved with such networks. Less ambitiously than aiming for a complete understanding, we can search for characteristics that indicate good generalisation.

We propose that the structural characteristics of the Representation Manifold (RM) of trained networks contain key characteristics that indicate whether a model will generalise or not to downstream tasks.

We measure the characteristics by applying sequentially larger local alterations to the input data and then tracking the movement of each datapoint on the RM. We then combine these measurements into one metric, the *Representation Manifold Quality Metric (RMQM)*, where larger values indicate larger and less variable step sizes on the RM. With this, we show:

- Self-supervised methods learn an RM where alterations lead to large but constant size changes, indicating a smoother RM than fully supervised methods.
- RMQM correlates positively with performance on downstream tasks, thus, we can use it to predict the downstream task performance of a trained model, with a correlation coefficient of 0.75.

RM CHARACTERISTICS

$$D(\phi, A) = \frac{1}{NJ}\sum_{i}^{N}\sum_{j}^{J}\|\phi_{i,0} - \phi_{i,j}\|_2$$

$$D(\phi_i, A)_{RC} = \frac{1}{J}\sum_{j=1}^{J}\left|\frac{d(\phi_{i,0}, \phi_{i,j}) - d(\phi_{i,0}, \phi_{i,j-1})}{d(\phi_{i,0}, \phi_{i,j})}\right|$$

$$P(\phi_i, A)_{RC} = \frac{1}{J}\sum_{j=2}^{J}\left|\frac{d(\phi_{i,j-1}, \phi_{i,j}) - d(\phi_{i,j-1}, \phi_{i,j-2})}{d(\phi_{i,j-1}, \phi_{i,j})}\right|$$

LOCAL ALTERATIONS

White noise injection:
We increase epsilon from 0 to 100 to increase the alteration strength in J steps.

$$x_{ij} = [x_i + a_j]_{clip} \quad a_j \sim \mathcal{N}(0, \epsilon_j^2)$$

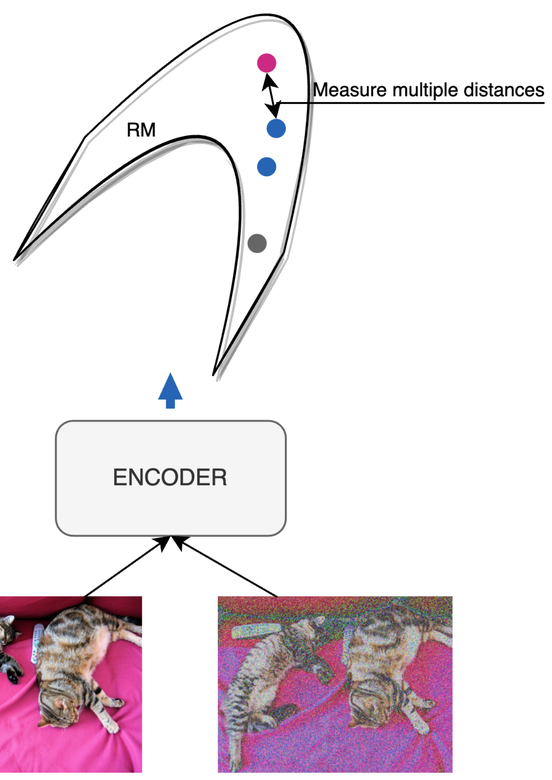
PGD Attacks:
We increase the amount of PGD steps to increase the alteration strength.

$$x_{i,j} = [x_{i,j-1} + \epsilon_{FSGM} \cdot (\nabla_{x_{i,j-1}} \mathcal{L}(\theta, x_{i,j-1}, y))]_{clip}$$

RMQM

RMQM is designed to yield large values for relatively smooth RMs with relatively large sensitivity to changes in the input.

$$RMQM = \ln\left(1 + D + D_{PC}^{-1} + P_{PC}^{-1}\right)$$



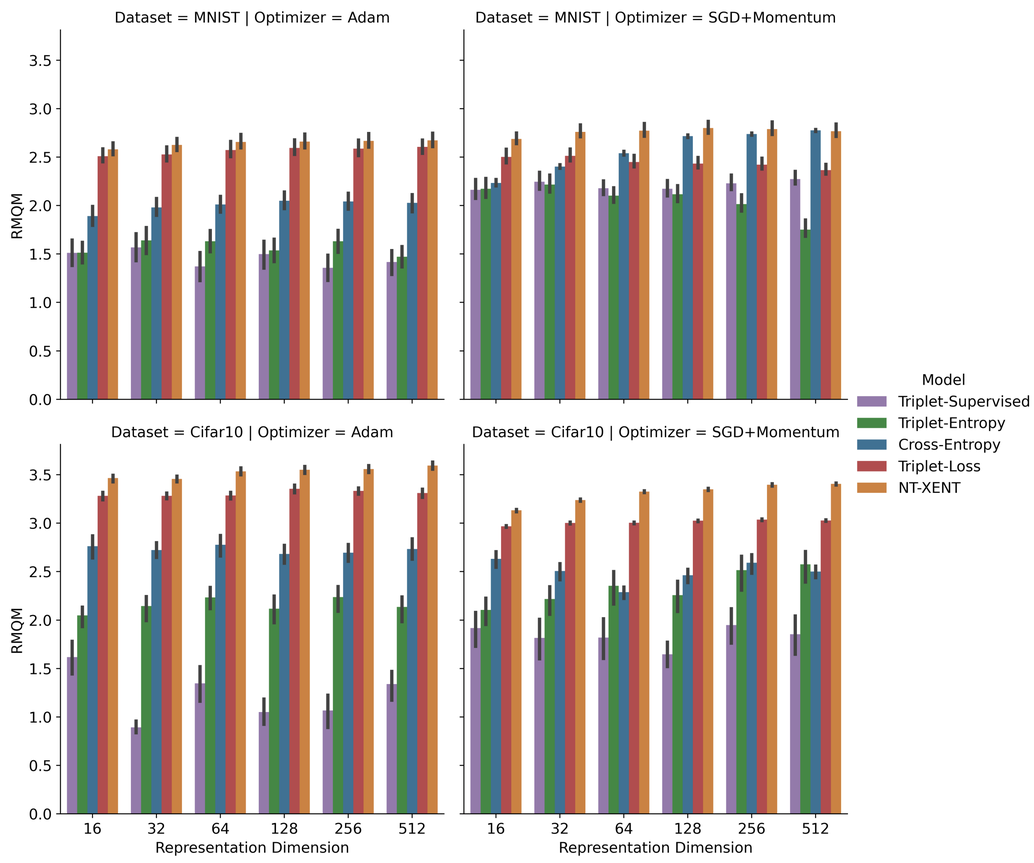


Table 1. The average distance each points moves relative to the original point for both the white noise injection and PGD Attack alterations for the MNIST encoders. The results are averaged over both the embedding dimension of an encoder and the optimiser used.

METHOD	NOISE	PGD
CROSS-ENTROPY	0.33±0.15	0.21±0.08
TRIPLET-ENTROPY	0.73±0.11	0.33±0.08
TRIPLET-SUPERVISED	0.77±0.06	0.38±0.10
TRIPLET-SS	0.93±0.13	0.66±0.17
NT-XENT	1.01±0.04	0.68±0.10

Table 2. The average change in the distance each point moves relative to the original point, compared against the previous alteration's distance. Results are calculated for both the white noise injection and PGD Attack alterations for the CIFAR-10 encoders. The results are averaged over both the embedding dimension of an encoder and the optimiser used.

METHOD	NOISE	PGD
CROSS-ENTROPY	0.11±0.03	0.34±0.04
TRIPLET-ENTROPY	0.18±0.04	0.89±0.35
TRIPLET-SUPERVISED	0.86±0.85	2.97±1.61
TRIPLET-SS	0.06±0.01	0.06±0.04
NT-XENT	0.04±0.01	0.10±0.01

IDEAL REPRESENTAION MANIFOLD

When vector search is the downstream task performance (KNN), an encoder that learned an RM with smooth structure and large displacements will tend to perform well on downstream search tasks.

One can interpret this as defining a new meaning for network robustness as well, where the model is not robust if there are small displacements in the output given changes in the input, but rather that it is robust if these displacements move with constant step sizes.

CONCLUSION

We show that self-supervised learning methods learn RMs in which motion in any direction on the surface will result in relatively large displacements. However, these displacements are relatively similar no matter where or in what direction a step is taken.

To identify RM characteristics related to good downstream task performance, we combine our measurements into a single metric, the Representation Manifold Quality Metric (RMQM). We find RMQM positively correlates with downstream task performance, and indicates that RM characteristics are a strong predictor for downstream task performance.

ByteFuse