Shape analysis in PET Images using Convolutional Neural Nets: Limitations of Standard Architectures

I. KLYUZHIN [1,2], A. RAHMIM [1,2]
1 University of British Columbia, Vancouver, Canada
2 BC Cancer Research Centre, Vancouver, Canada

JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

INTRODUCTION

Shape properties of tumors in PET, CT and MR images have been found to be significant predictors of disease progression and efficacy of treatment (1,2).

Thus, shape features have become an important component in radiomics-based pipelines (3).

On the other hand, the use of convolutional neural networks (CNNs) for image-based clinical tasks is gaining popularity. Recent work has determined that ImageNet-trained CNNs are biased towards texture (4). This implies that in medical image analysis, CNNs may implicitly under-utilize shape information.

AIM

To test the ability of practical CNN architectures to explicitly "learn" standardized radiomic shape features, in comparison to intensity and texture features.

To this end, we train CNNs to predict the values of radiomic features for synthetic PET images of tumors.

METHOD

Image data and radiomic features

- 5000 synthetic PET images of tumors (64x64x64 voxels) and their binary masks were generated (Figure 1) using a stochastic region growth algorithm and Perlin pattern generator.
- Radiomic features were computed using the SERA library (5).
- Shape features were computed from the binary lesion masks, while intensity and texture features were computed using voxel intensities inside the lesions.

Neural net architectures

- A series of standard "convolution-nonlinearity-pooling" (CNP) network architectures were tested, as well as several state-of-the-art (SOTA) networks pre-trained on ImageNet.
- Standard 3D CNN architectures were tested with 3, 5, 7 and 9 convolutional layers. SOTA networks included: MobileNetV2, Xception, NASNetMobile, DenseNet201; only the final regression layer was trained.
- The inputs were the intensity images, and the targets were the corresponding radiomic feature values.
- 100 training epochs, batches of 32 images, Adagrad optimization (learning rate 0.01), mean absolute error loss function.
- 4000 images were used for testing, 500 for validation, and 500 for testing.

Analysis

- The agreement the between the CNN-predicted and explicitly-computed radiomic feature values was analysed.

RESULTS: STANDARD CNN ARCHITECTURES

- Prediction errors were measured as the range-normalized mean absolute error between the predicted and ground truth feature values.
- The highest prediction errors were measured for features that quantified the shape irregularity – solidity, extent, elongation, and compactness (Figure 2).
- Notably, these features were predicted least accurately with every CNN, i.e. regardless of the network depth.
- The lowest prediction errors were measured with size features (area, convex area, eq. diameter, perimeter), as well as the mean and max intensity values.
- The improvement in performance with added convolutional layers was either small or insignificant.

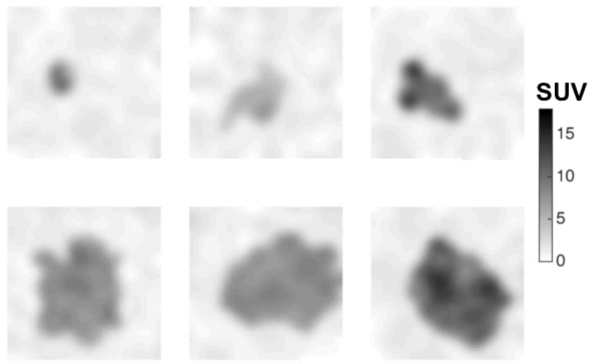
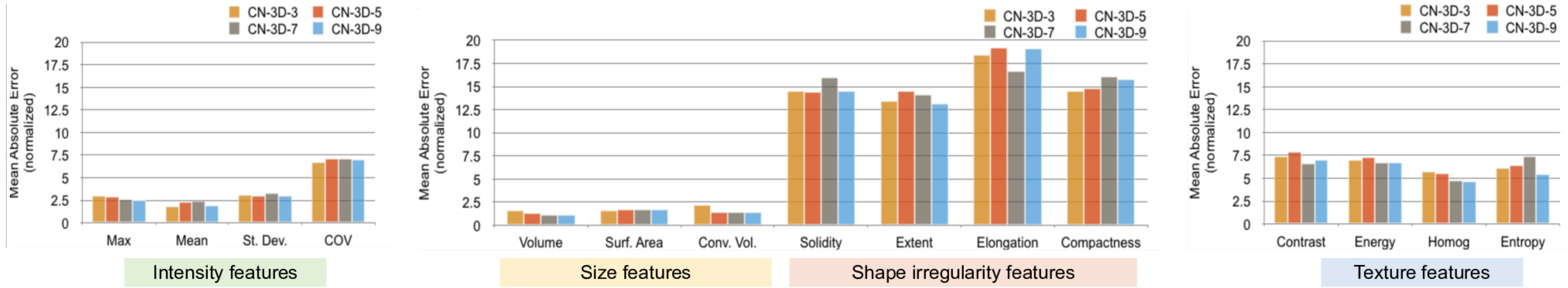


Figure 1: Examples of generated synthetic PET images of tumors, illustrating different tumor sizes, shapes and textures.

The pixel intensities were set to represent PET standardized uptake values (SUV).

Figure 2: Radiomic feature prediction errors with standard CNN architectures. The CN-3D-*X* abbreviations denote 3D CNP networks, *X* stands for the number of convolutional layers. Mean error values are plotted from 5 independent CNN training trials.



RESULTS: STATE OF THE ART NETWORKS

- With 2D SOTA networks, the highest prediction errors were found with shape irregularity features: solidity, extent, and eccentricity (Figure 3) – a similar finding to standard CNN architectures.
- Overall, all SOTA networks performed similarly across different features: a greater number of parameters or layers in the network did not result in lower prediction errors.
- The SOTA prediction errors were higher compared to standard CNNs.
- The scatter plots for MobileNetV2 (Figure 4) demonstrate that the measured prediction errors did not originate from a few significant outliers or biases.

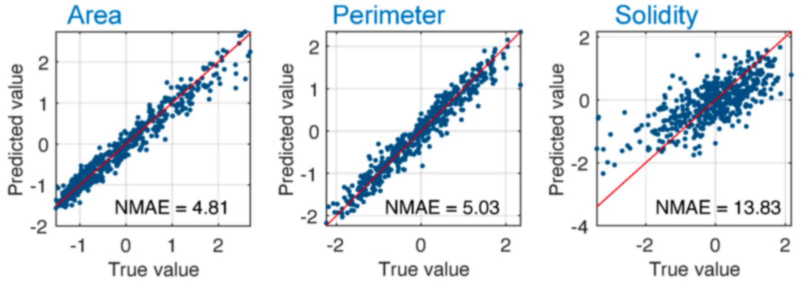
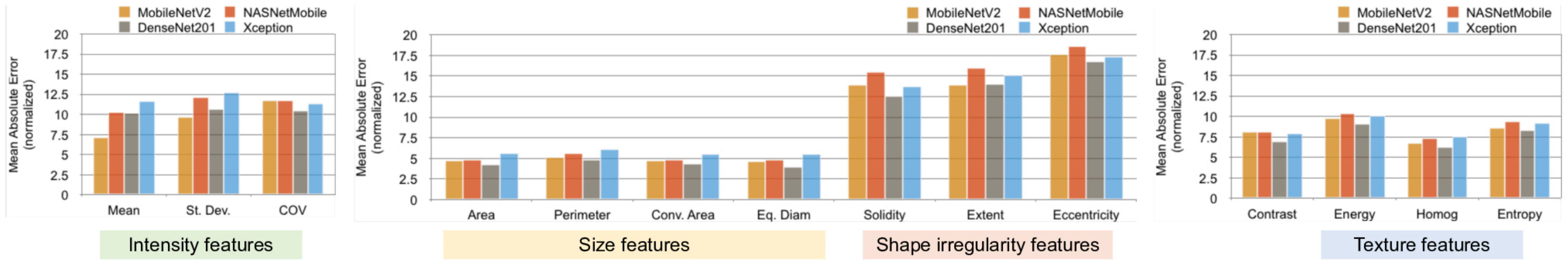


Figure 4: Predicted feature values plotted against true feature values for the MobileNetV2, NMAE = normalized mean absolute error.

Figure 3: Radiomic feature prediction errors with ImageNet-pretrained advanced networks. Mean error values are plotted from 3 independent CNN training trials.



CONCLUSIONS

- Standard CNN architectures and SOTA networks produce high prediction errors for shape irregularity features, such as solidity and extent. Size, intensity and texture features are captured more readily, i.e. with lower prediction errors.
- Deep learning models, particularly CNNs, may not be effective at capturing and leveraging shape lesion properties that have previously been associated with clinical outcomes.
- The use of explicit radiomics and traditional machine learning techniques may not be readily discarded in favour of CNNs when it comes to medical image analysis, as the strengths of these two approaches appear to be complementary.
- Future work will focus on the strategies to improve shape feature representations in CNNs, as well as extending the tests to more realistic and extensive images of tumors.

ACKNOWLEDGEMENTS

We gratefully acknowledge CIHR Grant OQI-137993 / NIH U01 CA190232.

REFERENCES

1. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. Abdom Radiol. 2019;44(6):1960–84.

2. Hsu C-Y, Wang C-W, Kuo C-C, Chen Y-H, Lan K-H, Cheng A-L, et al. Tumor compactness improves the preoperative volumetry-based prediction of the pathological complete response of rectal cancer after preoperative concurrent chemoradiotherapy. Oncotarget. 2017 Jan 31;8(5):7921–34.

3. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020 May;295(2):328–38.

4. Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. 7th Int Conf Learn Represent ICLR 2019. 2018 Nov 29;(c):1–22. ArXivID:1811.12231

5. Ashrafinia S. Quantitative nuclear medicine imaging using advanced image reconstruction and robotics. PhD Dissertation, Johns Hopkins University; 2019.

CONTACT INFORMATION

Ivan Klyuzhin, PhD, Email: iklyuzhin@bccrc.ca