

LOSSLESS COMPRESSION FOR VOLUMETRIC MEDICAL IMAGES USING DEEP NEURAL NETWORK WITH LOCAL SAMPLING

O. H. Nagoor, J. Whittle, J. Deng, B. Mora, M. W. Jones

Department of Computer Science, Swansea University, Swansea, UK

ABSTRACT

Data compression forms a central role in handling the bottleneck of data storage, transmission and processing. Lossless compression requires reducing the file size whilst maintaining bit-perfect decompression, which is the main target in medical applications. This paper presents a novel lossless compression method for 16-bit medical imaging volumes. The aim is to train a neural network (NN) as a 3D data predictor, which minimizes the differences with the original data values and to compress those residuals using arithmetic coding. We evaluate the compression performance of our proposed models to state-of-the-art lossless compression methods, which shows that our approach accomplishes a higher compression ratio in comparison to JPEG-LS, JPEG2000, JP3D, and HEVC and generalizes well.

Index Terms— Lossless Compression, 3D Predictor, Medical Image Compression, Volumetric Data Compression, Neural Network

1. INTRODUCTION

Medical imaging is used for clinical diagnosis. Precise medical imaging techniques have been developed where radiologists can acquire high quality and high resolution scans for clinical purposes. 3D medical imaging is often used for further diagnosis and precise pre-surgery planning. According to Diagnostic Imaging Dataset Statistical Release published by NHS, between September 2018 to September 2019, there were over 45 million medical images acquired for clinical use including 5.8M CT scans and 3.7M MRI scans. Data storage for a large amount of medical images poses a great challenge. Especially for clinical purposes, artefacts that introduced by lossy compression introducing could result in misleading diagnosis and unfavourable treatment [1].

Lossless compression standards are classified into two main types – image encoders and volumetric encoders. Standard image encoders include JPEG2000 [2], Lossless and Near-Lossless Compression of Continuous-Tone Still Images (JPEG-LS) [3], Context-based Adaptive Lossless Image Codec (CALIC) [4], and Minimum Rate Predictor (MRP) [5]. In contrast to 2D image encoders, volumetric encoders enhance the compression ratio by applying a reduction in a

higher dimensional context, such as: JPEG2000 Part 10 Extensions (JP3D) [6], High-Efficiency Video Coding (HEVC) [7], 3D-CALIC [8], M-CALIC [9], and 3D Minimum Rate Predictor (3D-MRP) [10]. Lossless compression methods also have challenges such as relying on hand-crafted linear transformation and codecs have limitation in representing non-linear correlations. Recently, state-of-the-art deep neural networks are demonstrated as feasible to construct both lossy and lossless image compression which also achieve higher compression ratio compared to classic linear methods.

State-of-the-art deep learning approaches address lossy reduction to assist purposes such as dimensionality reduction (autoencoders) [11], super-resolution images or video reconstruction [12, 13, 14], estimating pixel likelihood (autoregressive) [15, 16], and generative compression [17, 18]. Less attention has been made to address the lossless performance using NNs. The current deep learning literature for lossless compression usually combine a density estimator model with an arithmetic coder or Asymmetric Numeral System (ANS). The density estimator can be categorised into various types, namely, fully connected NN [19], Recurrent Neural Network (LSTM/GRU) such as DeepZip [20], a recursive bits-back coding with hierarchical latent variables known as Bit-Swap [21], and a parallelizable hierarchical probabilistic model that learned both image distribution and its auxiliary representations L3C [22]. Our proposed model for lossless compression has a fully connected deep NN as its data estimator, which should be faster to train than a Recurrent Network in DeepZip and estimate higher dimensional data (3D medical images) compared to [19], which compresses only 2D scans. Additionally, our proposed 3D sampling scheme allows the model to generalise well to unseen samples.

Our main contributions are: (a) A novel 3D predictor model using a neural network that achieves lossless compression for volumetric medical images. (b) A computationally efficient model that achieves a higher compression ratio when compared to state-of-the-art lossless compression methods. (c) Demonstrate the robustness and generalization of our proposed models experimentally on many datasets for higher dynamic range (16 bit-depths).

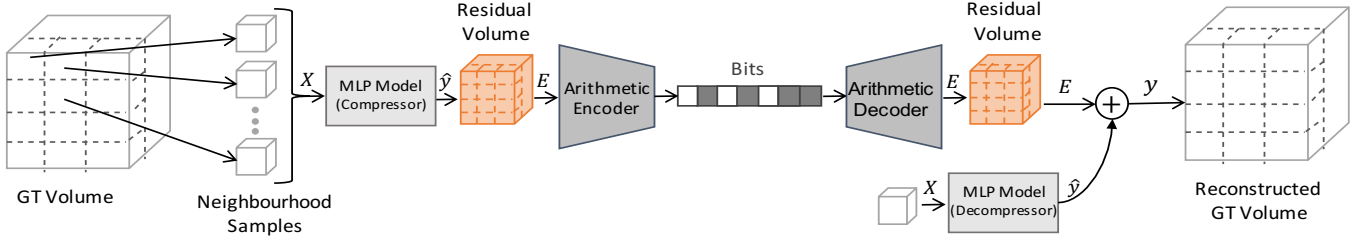


Fig. 1. An overview of the proposed lossless compression method.

2. PROPOSED METHOD

We propose (Fig. 1) a lossless compression approach using neural network specifically for volumetric medical images, where the data compression is formulated as a sequential prediction problem. Our approach consists of: the data prediction module which also called compressor, the entropy coding using arithmetic coder, and the data recovery module namely decompressor. Both compressor and decompressor use one neural network model where the architectures are the same and the weights are shared. The model learns the projection function to predict the target voxel given a sequence of samples from its neighbourhood (Fig. 2). In order to achieve high compression ratio for the arithmetic encoder, we train the neural network to minimize the difference between the prediction and the ground-truth. The regression problem can be solved by learning a mapping function f that predict the output y from an input sequence X through the back-propagation process given a training dataset. The hypothesis is that the prediction is highly correlated with the local appearance and geometric structure of its neighbours. Formally, given a data distribution defined over $X \in R^N$, where X contains input samples from the same distribution (e.g. $X = \{x_1, x_2, \dots, x_n\}$ forms a 1D vector of immediately neighboring voxel-intensities), we learn a differentiable mapping function $\hat{y} = f(X)$ that maps the vector X to a predicted value \hat{y} to minimize the differences with the ground truth voxel value y , where $f(X)$ is represented using a neural network model. Therefore, the residual or prediction error E is calculated as follows:

$$E = y - \hat{y} \quad (1)$$

The residual error E is then encoded using an arithmetic coder and transmitted in a lower bit-rate. The better the model performs in approximating the data distribution, the smaller the residual gets and the lower code-length the coder produces. To fully recover the original data from the compressed representation, the bit-stream is decompressed first, and then the residual values E are added to the prediction values \hat{y} (Fig. 1).

2.1. Local Sampling

The input sequence forms a crucial role in learning the mapping function of the data distribution. We experimented with differing sampling neighbourhoods and sampling strategy in order to find optimal compression. There is a trade-off between the amount of information presented to the model and the computation cost. Many image-based codecs use the immediate neighbouring pixels (i.e. four previous pixels) to predict the current pixel location. In 3D, neighbours from previous slices are included. These neighbouring sequences allow the encoder to exploit the spatial correlation and redundancy in the 2D plane as well as within the inter-frame region.

Training a model on samples uniformly selected across multiple volumes is problematic, as such a high dynamic searching space does not allow the model to accomplish a maximum compression performance. Therefore, in our proposed work, we reduce the sampling space to one volume with the assumption that the data distribution of the human body across the volumes would share some structural similarity and common feature representation since all scans share the same resolution across width and height and are captured by the same hospital with the same scan parameters (for training set and test set 1). Instead of sampling randomly from the whole volume (uniform voxel sampling) or biasing our sampling towards part of it (3D Gaussian voxel sampling), we extract multiple complete 2D slices across the volume z-axis with a fixed stride (i.e. ten slices from one volume). Then, for each voxel in the ten slices, we extract the 3D neighbouring voxels. We introduce two different shapes of the neighbouring blocks; namely, 3D cube neighbouring sequence and 3D pyramid neighbouring sequence (See Fig. 2). In both types, the sequences never include voxels from future slices. All volume values are normalized to the range $[-1, 1]$ and the volume is padded, as determined by the block size, by its minimum voxel value. Padding the volume is crucial in order to include the edge and corner cases in training. All the 3D sequences will be flattened to 1D vectors and randomly shuffled before inputting them to the predictor models. The reason for selecting such sampling schemes is to provide better coverage of the data to benefit the compression ratio. We find that training the model on whole slices with every

voxel within the selected slices being available during training leads to an improvement in final compression ratio on all slices including those not trained on.

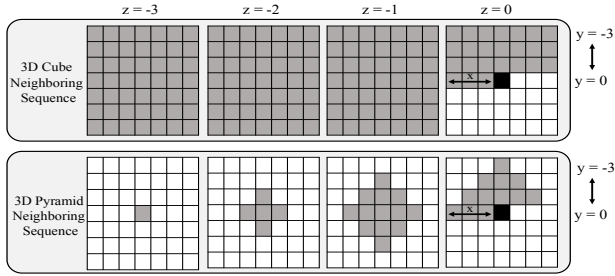


Fig. 2. Two neighbourhoods used for prediction. $z=0$ represents the current slice, the target voxel for prediction is black, grey voxels are used in the input sequence.

2.2. Transforming Function

| Layer | Number of Neurons | Activation Function Used |
|-----------------|-------------------|--------------------------|
| Fully Connected | 1024 | LeakyRelU |
| Fully Connected | 512 | LeakyRelU |
| Fully Connected | 256 | LeakyRelU |
| Fully Connected | 128 | LeakyRelU |
| Output | 1 | Linear |

Table 1. The proposed neural network architecture.

Multi-layer Perception (MLP) is used to build the sequence prediction models, which consists of an input layer, 4 fully connected hidden layers with non-linear activation functions and followed by a linear output layer as output. The parameter settings of individual layers are given in Table 1. Inspired by [23], to train the neural network, we minimize a joint loss L_{Joint} (Eq. 2) of the sum of Mean Absolute Error (MAE) (Eq. 3) and the Pearson Correlation Coefficient (PCC) (Eq. 4)

$$L_{Joint} = MAE + \lambda(1 - |PCC|) \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n |y - \hat{y}|}{n} \quad (3)$$

$$PCC = \frac{cov(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}} \quad (4)$$

where, cov is the co-variance of ground-truth and the prediction, σ_y and $\sigma_{\hat{y}}$ are the standard deviation of y and \hat{y} respectively. PCC measures the linear correlation between two variables y and \hat{y} . Empirically, we found that joining those two losses improves the accuracy and stabilizes the training process.

3. RESULT & DISCUSSION

All the datasets used in this paper consist of a set of DICOM files, which represent CT scans for patient’s entire trunk

stored in 16-bit grayscale images. Our models were trained on one training set however the evaluation were conducted on two different test sets. Both training set and testing set 1 are a private dataset from the same data source – generated by the same hospital and with similar scanning parameters such as the slice thickness, and spacing between slices but with some variation in the pixel spacing $\in [0.488, 0.5, 0.635, 0.703]$. All volume slices have 512×512 resolution but it varies in the number of frames (volume depth), which is in the range of 750 to 1120 frames. The maximum value is 3071 and the minimum value is -1024. The training set consist of ten slices out of 840 slices extracted from one volume (patient 40) with pixel spacing $[0.625, 0.625]$ and slice thickness 0.625. The two proposed NN models were evaluated on test set 1, which consist of 42 volumes. On the other hand, test set 2 contains two volumes from a publicly available dataset provided by The Cancer Imaging Archive (TCIA) [24]. The DICOM dataset is CT scans of a human lung for patients who suffer from non-small cell lung cancer. The datasets characteristics are presented in table 3.

To select the block size with best compression performance, we experimentally applied different neighbouring sizes for the 3D cube input sequence including $3 \times 3 \times 3$, $5 \times 5 \times 5$, $7 \times 7 \times 7$, $9 \times 9 \times 9$, and $11 \times 11 \times 11$. Figure 2 illustrates an example of each input sequence type. In the given examples, $z=0$ represents the current slice, the black voxel refers to the target voxel that needs to be predicted, and the grey voxels are the input sequences while the white voxels are the ones that will be masked because they are unknown information during decompression. In the given 3D cube example, the block size is equal to $7 \times 7 \times 7$, which implies that the maximum distance from target voxel in each dimension (x, y, z) is 3. However, in the 3D pyramid neighbouring example, the distance to the target voxel is decreasing in both x and y axes at each increasing z step. For instance, (at $z=0$, the block size= 7×7), and (at $z=-1$, the block size= 5×5), etc. Based on several experiments, we choose ($11 \times 11 \times 11$) as the input sequence block for model 1 as it produces the best compression rate. However, for model 2, the 3D pyramid sequences with (13×13 , 9×9 , 5×5 , 1×1) sequence size are used. The rationale for choosing a pyramid is that it reduces the number of voxels in each training sample (and consequently storage size and training time) but still retains the possibility to integrate information from previous slices into the training.

For the neural network, we set the parameters $\lambda = 1$, which weights the contribution of the two losses to be the same. We optimised the neural network using Stochastic Gradient Descent (SGD) with momentum $\beta = 0.9$. Both models have a good stability even without applying batch normalization, weights L2 regularization or dropout. The training hyperparameters of the two models are illustrated in table 2. The hyperparameters for model 1 are with batch size of 256, learning rate of 0.0002. However, for model 2, the batch size of 32 and learning rate of 3×10^{-5} are used.

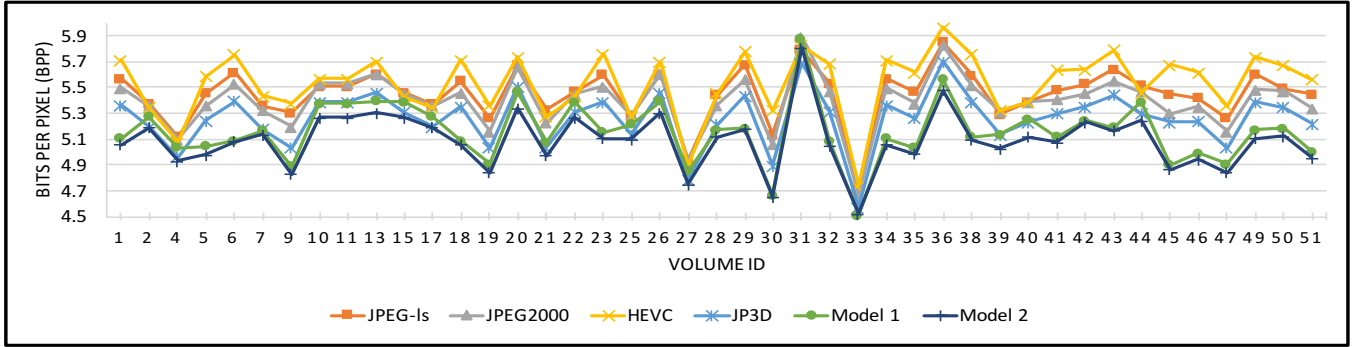


Fig. 3. Comparing the compression ratio in bpp for the proposed models with the state-of-the-art lossless compression methods over 16-bits volumes on test set 1.

| Model ID | Sampling Space | Shapes of the input Neighbouring Block | Hyper Parameters |
|----------|--|--|---|
| 1 | All samples were generated from 10 slices extracted from one volume (patient 40) | 3D Cube 11x11x11 input sequence | Batch size = 256, learning rate = 0.0002, no L2 regularization, no dropout and no batch normalization |
| 2 | All samples were generated from 10 slices extracted from one volume (patient 40) | 3D pyramid input sequence 13x13, 9x9, 5x5, 1x1 | Batch size = 32, learning rate = 3e-5, no L2 regularization, no dropout and no batch normalization |

Table 2. Illustrating the neural network training specifications for the two proposed models.

| Patient ID | Resolution | Frames | Pixel Spacing | Slice Thickness |
|--------------|------------|--------|----------------------|-----------------|
| CT_Lung_R004 | 512 x 512 | 68 | [0.830078, 0.830078] | 5.0 |
| CT_Lung_R013 | 512 x 512 | 67 | [0.623047, 0.623047] | 5.0 |

Table 3. Overview of test set 2 information composed of 16 bit-depth medical images [25] [26]

We evaluated the compression performance in bits-per-pixel (bpp) of the proposed neural network models in comparison to the state-of-the-art lossless compression methods including JPEG-LS, JPEG2000, JP3D and HEVC (HM-SCC-extensions-4998) using the lossless configuration with main-RExt profile available in [7]. Figure 3, illustrates the compression rate in bpp on test set 1 compressed by the state-of-the-art lossless methods and our two proposed neural network models. The results indicate that the proposed predictor models achieve the best compression ratio on test set 1 in comparison to the existing methods. Additionally, it is clear that the methods using 3D contents (two proposed models, and JP3D) gained smaller bit-rate than the ones using 2D content only (JPEG2000 and JPEG-LS). Among the different compression approaches, our 3D data predictor model (Model 2) achieves the best compression ratio.

We also evaluated the generalization ability of our models on a completely different data distribution (CT of lung cancer). Our models were not trained on these volumetric medical images, however, it can again achieve close compression ratio to other methods: *Dataset Lung_R004*, JPEG-LS 5.937bpp, JPEG2000 6.014bpp, **HEVC 5.739bpp**, JP3D 5.967bpp, Model 1 6.664bpp, Model 2 6.715bpp and *Lung_R013*, JPEG-LS 5.747bpp, **JPEG2000 5.539bpp**, HEVC 5.835bpp, JP3D 5.623bpp, Model 1 5.959bpp, Model

2 5.847bpp. This is achieved even though the scanning parameters of the test set 2 differs totally from the training set and test set 1. The slice thickness is 5mm while in training set and test set 1, the thickness is .625 mm, which influences the 3D neighbouring quality learned by our models. Since our neural network models were trained to learn the explicit relation of neighbouring voxels in a specific resolution (i.e. training set), different volume resolution (i.e. test set 2) will influence the performance of the models. However, our model is generalized to gain better compression if the data provided is consistent and has similar structure to the training set (i.e. test set 1 has similar slice thickness but variation in pixel spacing). We tested this by resampling test set 2 to pixel spacing of 0.625 and compressing it. On that test, our models performed better than the other methods (*Lung_R004*, JPEG-LS 5.46bpp, JPEG2000 5.24bpp, JP3D 5.2bpp, Model 1 4.92bpp, **Model 2 4.9bpp** and *Lung_R013*, JPEG-LS 5.698bpp, JPEG2000 5.485bpp, JP3D 5.375bpp, **Model 1 5.237bpp**, Model 2 5.238bpp).

4. CONCLUSION

In this paper, we proposed a novel lossless compression system using neural network for volumetric medical image. Two localized sampling methods were introduced and evaluated on real volumetric medical imaging datasets. The comparison study shows that our method outperforms the standard lossless compression methods. It also suggests that the proposed method is feasible to generalize to unseen dataset while retains satisfactory performance.

5. REFERENCES

- [1] D. Koff and H. Shulman, "An overview of digital compression of medical images: Can we use lossy image compression in radiology?," *CAR Journal*, vol. 57, pp. 211–7, 11 2006.
- [2] D. Taubman and M. Marcellin, *JPEG2000 image compression fundamentals, standards and practice: image compression fundamentals, standards and practice*, vol. 642, Springer Science & Business Media, 2012.
- [3] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The loco-i lossless image compression algorithm: principles and standardization into jpeg-ls," *IEEE Trans. Image Process*, vol. 9, no. 8, pp. 1309–1324, Aug 2000.
- [4] X. Wu and N. Memon, "Calic-a context based adaptive lossless image codec," *1996 IEEE ICASSP*, vol. 4, pp. 1890–1893 vol. 4, May 1996.
- [5] I. Matsuda, H. Mori, and S. Itoh, "Lossless coding of still images using minimum-rate predictors," *Proceedings ICIP 2000*, vol. 1, pp. 132–135 vol.1, Sep. 2000.
- [6] P. Schelkens, A. Munteanu, A. Tzannes, and C. Brislawn, "Jpeg2000. part 10. volumetric data encoding," *IEEE ISCAS 2006*, pp. 4 pp.–3877, May 2006.
- [7] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, J. Sole, and J. Xu, "Overview of the range extensions for the hevc standard: Tools, profiles, and performance," *IEEE TCSVT*, 2016.
- [8] X. Wu and N. Memon, "Context-based lossless interband compression-extending calic," *IEEE Trans. Image Process*, vol. 9, no. 6, pp. 994–1001, June 2000.
- [9] E. Magli, G. Olmo, and E. Quacchio, "Optimized on-board lossless and near-lossless compression of hyperspectral data using calic," *IEEE GRSL*, vol. 1, no. 1, pp. 21–25, Jan 2004.
- [10] L. F. R. Lucas, N. M. M. Rodrigues, L. A. da Silva Cruz, and S. M. M. de Faria, "Lossless compression of medical images using 3-d predictors," *IEEE TMI*, vol. 36, no. 11, pp. 2250–2260, Nov 2017.
- [11] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," *ICLR*, vol. abs/1703.00395, 2017.
- [12] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 CVPR*, June 2016, pp. 1646–1654.
- [13] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," *2017 CVPR*, pp. 5835–5843, 2017.
- [14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 CVPR*, July 2017, pp. 105–114.
- [15] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, "Full resolution image compression with recurrent neural networks," in *2017 CVPR*, July 2017, pp. 5435–5443.
- [16] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *ICML*, 2016.
- [17] S. Santurkar, D. Budden, and N. Shavit, "Generative compression," in *PCS 2018*, June 2018, pp. 258–262.
- [18] Y. Liu, Y. Wang, L. Deng, F. Wang, F. Liu, Y. Lu, and S. Li, "A novel in situ compression method for cfd data based on generative adversarial network," *Journal of Visualization*, vol. 22, no. 1, pp. 95–108, Feb 2019.
- [19] M. U. Ayoobkhan, E. Chikkannan, and K. Ramakrishnan, "Feed-forward neural network-based predictive image coding for medical image compression," *AJSE*, vol. 43, no. 8, pp. 4239–4247, Aug 2018.
- [20] M. Goyal, K. Tatwawadi, S. Chandak, and I. Ochoa, "Deepzip: Lossless data compression using recurrent neural networks," in *2019 DCC*, March 2019.
- [21] F. H. Kingma, P. Abbeel, and J. Ho, "Bit-swap: Recursive bits-back coding for lossless compression with hierarchical latent variables," in *ICML*, 2019.
- [22] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Practical full resolution learned lossless image compression," in *Proceedings of the CVPR*, 2019.
- [23] J. Whittle and M. W. Jones, "A deep learning approach to no-reference image quality assessment for monte carlo rendered images," *CGVC 2018*, pp. 23–31, Sept 2018.
- [24] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (tcia): Maintaining and operating a public information repository," *JDI*, vol. 26, pp. 1045–1057, Dec 2013.
- [25] B. Zhao, L. P. James, C. S. Moskowitz, P. Guo, M. S. Ginsberg, R. A. Lefkowitz, Y. Qin, G. J. Riely, M. G. Kris, and L. H. Schwartz, "Evaluating variability in tumor measurements from same-day repeat ct scans of patients with non-small cell lung cancer," *Radiology*, 2009.
- [26] B. Zhao, L. H. Schwartz, and M. G. Kris, "Data from rider lung ct. the cancer imaging archive," *TCIA*, 2015.