ASSESSING THE IMPACT OF THE DECEIVED NON LOCAL MEANS FILTER AS A PREPROCESSING STAGE IN A CONVOLUTIONAL NEURAL NETWORK BASED APPROACH FOR AGE ESTIMATION USING DIGITAL HAND X-RAY IMAGES

S. Calderon[⋆], *F. Fallas*[†], *M. Zumbado*[‡], *P. N. Tyrrell*[±], *H. Stark*[⋄], *Z. Emersic*[§], *B. Meden*[≀], *M. Solis*[+]

[⋆‡] School of Computing, Instituto Tecnológico de Costa Rica
[†] Atlantic Campus, Computing, Universidad de Costa Rica
[±⋄] Departments of Medical Imaging and Statistical Sciences, University of Toronto
[§≀] Faculty of Computer & Information Science, University of Ljubljana
[+] School of Business Administration, Instituto Tecnológico de Costa Rica
Email: [⋆]sacalderon@itcr.ac.cr, [†]fabian.fallasmoya@ucr.ac.cr, [‡]manzumbado@ic-itcr.ac.cr,
[±]pascal.tyrrell@utoronto.ca, [⋄]hershel.stark@mail.utoronto.ca,
[§]ziga.emersic@fri.uni-lj.si, [≀]blaz.meden@fri.uni-lj.si, [+]marsolis@itcr.ac.cr

ABSTRACT

In this work we analyze the impact of denoising, contrast and edge enhancement using the Deceived Non Local Means (DNLM) filter in a Convolutional Neural Network (CNN) based approach for age estimation using digital X-ray images from hands. The DNLM filter presents two parameters which control edge enhancement and denoising. Increasing levels were tested to assess the impact of both contrast enhancement and denoising in the CNN based model regression accuracy. Results obtained showed that contrast enhancement was important for preprocessing in a CNN based approach, given a statistically significant 42% lower root mean squared error, with comparable to previous state of the art results, using larger publicly available dataset. The obtained results suggest that both image enhancement and denoising can significantly improve results in a CNN based model.

***Index Terms*—** x-rays, neural networks, image processing, signal processing, convolution

1. INTRODUCTION

Image processing in biology, microbiology, medicine, and related fields is becoming more widespread for medical applications that make use of digital imaging systems such as computed tomography and magnetic resonance imaging [1]. For instance in [2], authors present a cell tracking framework aimed at measuring glioblastoma tissue response to chemotherapy. In [3], a lung nodule detection system is described as one of the first biomedical image analysis system using convolutional neural networks (CNNs) that includes preprocessing as a first stage of its pipeline to enhance the overall system accuracy. However, detailed impact assessments of preprocessing algorithms and its parameters in biomedical image applications have not yet been fully addressed in the literature. Furthermore, continual deep and extensive analysis of the impact of preprocessing techniques in complex computer vision systems is important as models and preprocessing techniques develop. This paper aims to analyze the impact of a preprocessing stage in a CNN based model for a biomedical application that estimates age from digital X-ray images, using the filter proposed in [4] with different parameters. Sub-section 1.1 details the specific application domain of age estimation using bone X-ray images, and Sub-section 1.2 explores recent work done in image preprocessing impact assessment.

1.1. Age estimation using digital X-ray images from hand bones

In pediatrics, estimation of skeletal maturity using X-ray images is often performed by physicians interested in comparing patient bone age with their chronological age. The radiological examination uses the left hand X-ray image using either the Greulich and Pyle or the Tanner-Whitehouse methods [5]. Such comparisons help to diagnose and observe the effects of endocrine and metabolic disorders [6].

The usage of machine learning to estimate bone age using digital X-ray images has been explored by others, with recent deep learning based techniques being the most successful approaches [6, 5]. In [6] a three-stage pipeline model was proposed, consisting in a segmentation, denoising, and enhancing stages and later implements a CNN with a fully connected layer at its end for regression. Specifically, a GoogLeNet CNN model was implemented yielding a lowest Root Mean Squared Error (RMSE) reported of 0.82 years for a data set of around 8325 samples, ranging ages from 5 years and up [7].

Similar work was developed in [5], where a comparison of different pre-trained popular models as OverFeat, GoogLeNet and OxfordNet was performed. The authors proposed the usage of a deformation layer in order to enhance model robustness to input image distortions. The lowest Mean Absolute Error (MAE) reported was 0.79. The dataset used in this work was publicly released in [7], consisting in 1400 samples, from subjects of different races, with ages ranging from 1 to 18 years. As mentioned earlier in this section, in this work we propose to implement the usage of the DNLM filter as a preprocessing stage, and assess its impact in a CNN based approach, an aspect not addressed yet in the explored literature.

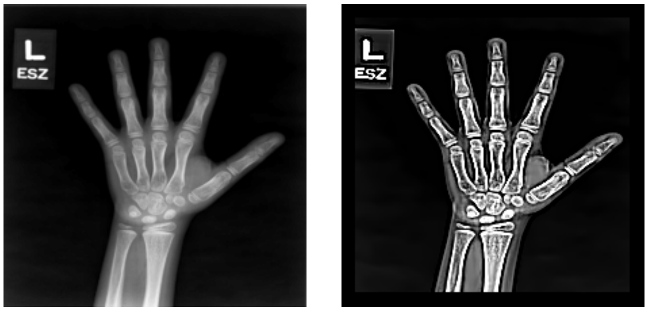
1.2. Preprocessing impact assessment in image based pattern recognition tasks

Research related to preprocessing impact assessment can be found in [8, 9, 10, 11]. In [10] image preprocessing using the median filter and the Non Local Means (NLM) filter for hand-crafted feature extraction based systems is assessed, using local binary patterns and histogram of gradients as feature extractors and a Support Vector Machine (SVM) for classification. Similarly, the work presented in [8] evaluates the impact of using a hybrid directional lifting technique for image denoising in a SVM classification algorithm. For the experimentation implemented in both papers, authors conclude that the preprocessing stage tested boosts classification precision.

Among the most important recent breakthroughs in image processing is the implementation of Convolutional Neural Networks (CNNs) since the publication of [12]. CNNs implement an overall modification of the classical pattern recognition pipeline, where feature extraction is automatically performed by the CNN model, dismissing hand crafted feature extraction and letting the model learn features from the data. The quality of such learned features depends on the amount of data and its variance, as over-fitting might occur in scenarios with few or not meaningful training samples. Thus, deep learning models often implement a data augmentation stage, where adding noise to image samples is common for such purposes [12]. Based on [12, 13, 14, 15] we can categorize data augmentation techniques in two major approaches: image transformation based (scale, translation and rotation modifications) and color manipulation based (lighting, contrast and noise manipulations) where authors argue that noise injection to input images acts as a powerful regularization technique[15]. Taking this concept further, adding noise to layer weights in a CNN has become a useful technique to avoid overfitting, an approach known as dropout [16].

Given the major paradigm shift implemented by CNN models, evaluating the impact of preprocessing techniques becomes appealing. In [9, 11], the impact of noise and denoising techniques is analyzed using a CNN based approach. More specifically, [11] shows that popular CNN models like VGG-16 and GoogLeNet are all negatively susceptible to noise and blur artificial degradations of training samples, but less sensitive to contrast degradation. With a different experimentation, authors in [9] compare CNN models for classification in popular data sets as MNIST, using noisy and denoised training samples. For noisy training samples, salt and pepper and Gaussian noise with different degradation intensities were tested, and as for denoised training samples, median and NLM filters were used, respectively. Perhaps counter intuitively, results showed that classification error is lower when using noisy training samples, compared to denoised samples. Results with salt and pepper noise showed this trend strongly, suggesting that using noisy training samples might work as a dropout behavior in the input layer of a CNN. However, in the aforementioned work, the impact of different denoising parameters of the techniques implemented and contrast in the input images of a CNN model was not addressed.

In this work, the objective is to extend the preprocessing impact assessment initially carried out in [9, 11], measuring the impact of enhancing contrast in the input images of a CNN based model. More specifically, we aim to explore the impact of edge and contrast enhancement and denoising of training samples using the Deceived Non Local Means (DNLM) filter, proposed in [4] and optimized in [17] in a real-world application: the estimation of bone age using digital hand X-ray images. Previous analysis of preprocessing impact using similar techniques to the DNLM were done in [18, 19], where an evaluation of the preprocessing technique impact over segmentation and tracking precision was performed, using hand-crafted based approaches for object segmentation and tracking, with good results. Typically, denoising algorithms as the NLM are evaluated using the peak signal to noise ratio or specific image quality metrics. We considered it is also important to assess the impact of using similar preprocessing techniques in a specific domain jointly with feature extraction and classification algorithms.



(a) Original image

(b) DNLM-IFFT output

Fig. 1. Output of the DNLM filter with a window size of 15×15, a neighborhood size of 3×3, $h = 12$, and $\lambda = 5$.

2. PROPOSED METHOD

This research proposes the usage of the DNLM-IFFT [17], a computational optimization of the DNLM as a preprocessing stage for a CNN based model for age estimation using X-ray images. The DNLM combines Unsharp Masking (USM) with a weighted averaging filter as the NLM [20], to obtain both noise reduction and, contrast and edge enhancement. The combination uses the USM image output, for instance with a simple linear implementation; $F_{\text{USM}} = U + \lambda L$, with λ controlling the amount of sharpening as input of the NLM filter, with the original input image U for weighting and L the laplacian of the input image. This avoids the ringing effect, an artifact often present in USM filtered images. The following equation formalizes the DeWAFF approach:

$$Y(p) = \left(\sum_{m \in \Omega} \psi_{\text{NLM}}(U, p, m) \right)^{-1} \left(\sum_{m \in \Omega} \psi_{\text{NLM}}(U, p, m) \, F_{\text{USM}} \right), \quad (1)$$

With p a specific pixel and Ω the current window. The NLM filter equation is defined as:

$$\psi_{\text{NLM}}(U, p, m) = \exp\left(-\frac{\| \overrightarrow{\eta}(\omega, m) - \overrightarrow{\eta}(\omega, p) \|^2}{h} \right), \quad (2)$$

with $\overrightarrow{\eta}(\omega, m)$ the vectorized neighborhood of ω pixels for the pixel m in U. The DNLM behavior is controlled by λ, defining the amount of sharpening and contrast enhancement, and h regulating image denoising and texture simplification. Figure 1 shows a comparison between one original X-ray image and a preprocessed image using the DNLM filter. In [17], an optimized implementation of the DNLM is proposed.

The DNLM-IFFT based processing of the input images serves as a previous stage of a CNN model for age estimation. Figure 2 shows the empirically tuned model details, along with the number of units and the activation functions per layer. The input to the network is a gray scaled image sized 256×256 pixels, and the output is the estimated age. We built two models (one per gender) and trained them separately to improve bone age predictions using gender annotations. The proposed architecture uses VGG16 as base model, empirically tuned, with weights initialized from a model trained for the ImageNet dataset [12].

We used 15 layers from the model, excluding its fully connected layer. Consequently, features are obtained from the VGG16 base model, which are passed to a customized fully connected network, known as the top model. The top model is trained separately using the extracted features as input, consisting in a 4-layer fully-connected regression network. As the objective for this model is to perform regression, we used a linear activation function for the output unit and the Mean Absolute Error (MAE) for error minimization.

The base VGG16 architecture [14] was selected due to the popularity and a good compromise between performance and complexity [21]. The top model was selected based on preliminary experiments that shown superior performance compared to other setups with 1, 2, 3, 5 or 6 layers. VGG16 [14] is a very deep CNN architecture. This means it should be able to capture complex patterns. However possible over-fitting needs to be considered in simpler cases [14] The main characteristic of the network is to use the smallest filter size capable of encoding directional information, i.e. 3×3 filters. These filters are then used consecutively in order to capture the same information as the larger filters, used with e.g. AlexNet [21]. This also means that significantly less parameters need to be estimated during training.

These 3×3 filter blocks are used with max-pooling layers which reduce activation maps dimensionality similar to other CNN architectures [21]. The convolutional part of the VGG model is in its original form followed by three fully-connected layers with $4,096$, $4,096$ and $1,000$ channels [14, 21].

The VGG16 component can be replaced with an arbitrary CNN architecture to further improve the results, in this paper we focused assessing preprocessing impact in a CNN based approach.

3. EXPERIMENTS AND RESULTS

The dataset consists in the publicly available digital images repository from left hands of both male and female subjects with ages ranging from 1 month to 228 months (19 years). The Radiological Society of North America (RSNA) made this dataset publicly available and was acquired from Stanford Children's Hospital and Colorado Children's Hospital.

Since λ and h are the most important parameters to be tuned for the DNLM filter, Analysis of Variance (ANOVA) was performed as defined by [22] to compare performance as done in [23] with statistical significance. Fixed window and neighborhood sizes were used, 15×15 and 3×3 respectively. We defined two factors, λ and h, with three and four levels, with 12 combinations or treatments, as Table 1 shows. We executed 10 replicas per treatment for a total of 120 runs. Table 1 shows the descriptive results for each treatment or parameter sets tested, the Root Mean Squared Error (RMSE) for the test data sets, given the usage of the MAE for training.

A two-way factorial arrangement on a randomized complete block analysis of variance was conducted on the influence of two independent variables (λ, h) on the ability to lower the RMSE. All effects were statistically significant at the .05 significance level. The main effect for λ type yielded an F ratio of $F(2, 99) = 251.1$, $p < .0001$, indicating a significant difference between level 0 ($\mu_{\lambda=0} = 20.2, \sigma_{\lambda=0} = 1.7$), level 2.5 ($\mu_{\lambda=2.5} = 17, \sigma_{\lambda=2.5} = 3.4$) and level 5 ($\mu_{\lambda=5} = 14.9, \sigma_{\lambda=5} = 3.7$). The main effect for h yielded an F ratio of $F(3, 99) = 247.5, p < .0001$, indicating that the effect for age was significant, level 0 ($\mu_{h=0} = 21.9, \sigma_{h=0} = 1.1$), level 8 ($\mu_{h=8} = 16.5, \sigma_{h=8} = 3.4$). However, there is not a positive improvement between level 12 ($\mu_{h=12} = 15.5, \sigma_{h=12} = 3.1$),

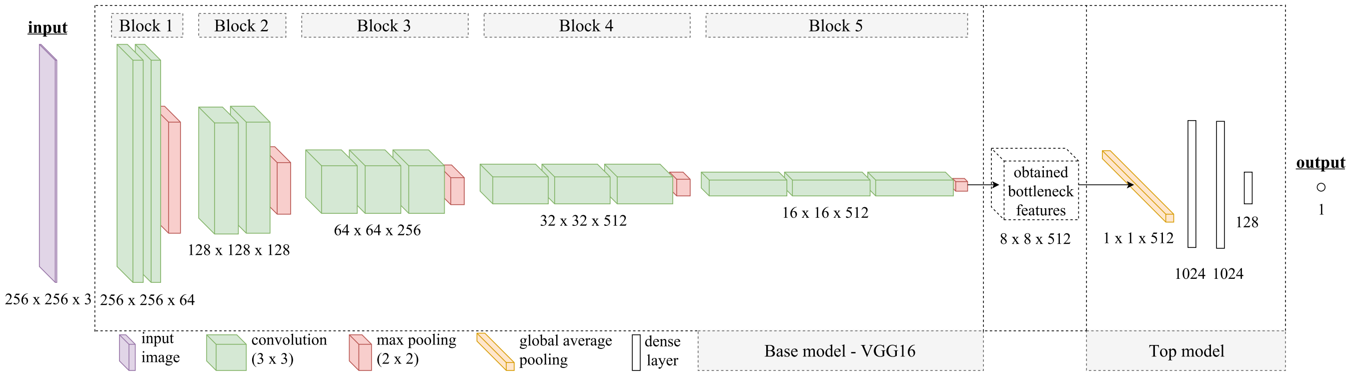


Fig. 2. Proposed CNN based pipeline: 256×256 gray scale pixel images as an input (the information is replicated for the three input channels). VGG16 as a base model to extract features, which are passed into the top model.

λ	h	**Avg.**	**Std.**	**Min.**
0	0	22.203	1.033	20.989
0	8	20.457	1.39	18.168
0	12	19.371	1.399	17.801
0	14	18.777	0.519	18.137
2.5	0	22.367	0.961	20.699
2.5	8	16.166	1.343	14.058
2.5	12	14.343	1.404	12.866
2.5	14	14.998	0.837	13.42
5	0	21.079	0.853	20.11
5	8	12.956	1.201	11.852
5	12	**12.803**	0.993	11.864
5	14	**12.889**	0.713	11.56

Table 1. An RMSE descriptive analysis in months, with average, standard deviation and minimum values of performing ten replicas for different combinations of the DNLM parameters λ and h. The highlighted values correspond to the best parameter combination of the DNLM, however the increase of $h = 12$ to $h = 14$ does not ensure a statistically significant improvement independently from λ, given the Tukey analysis.

and level 14 ($\mu_{h=14} = 15.6, \sigma_{h=14} = 2.6$). The interaction effect was significant between the two factors, $F(6, 99) = 21.0$, $p < .0001$.

The Tukey test for λ performed presenting a 95% confidence interval from -5.85 to -4.69 of RMSE variation changing $\lambda = 0$ to $\lambda = 5$. As for changing λ from $\lambda = 2.5$ to $\lambda = 5$ a significant impact is also obtained, ranging from -2.03 to -2.61 ($p < 0.05$), suggesting an important and statistically significant influence of contrast and edge enhancement for all the tested levels. However, the Tukey test for factor h reveals that we were not able to demonstrate an improvement of using a level of $h = 14$ over $h = 12$. Nevertheless, choosing a higher level of $h = 12$ over $h = 8$ or over $h = 0$ has an statistically significant impact in lowering the RMSE, demonstrating a positive impact in accuracy performing image denoising.

4. CONCLUSIONS AND FUTURE WORK

Our experiment results suggested that image preprocessing for a CNN based approach brings an impressive accuracy boost allowing us to obtain very similar results to state of the art, with a the lowest RMSE of 0.96 years and the best average MAE of 0.7908 years, using a larger dataset. The ANOVA allowed us to ensure that both denoising and contrast enhancement are important to improve the model accuracy with the proposed preprocessing step producing an improvement of 42% when using the best parameters for the DNLM-IFFT with statistical significance. We think it is important that researchers in the field consider different image enhancement approaches to improve accuracy in CNN based models for regression and classification.

As future work, testing different CNN models for age estimation like the more sophisticated approaches proposed [6, 5] using a preprocessing stage as the one suggested in this work, is still pending, and would likely further improve the overall system's accuracy. Also, data augmentation techniques based on similar preprocessing techniques are worth to explore.

Moreover, given the concluded importance of a preprocessing approach in a CNN based model in this work, we aim to incorporate the DNLM-IFFT or any similar preprocessing approach within the CNN model, in order to calibrate its parameters using the error gradient calculated in every epoch during the CNN training. Such approach would avoid the need to separately calibrate the DNLM parameters and optimize them for the given dataset.

5. ACKNOWLEDGEMENTS

This research was supported by the Kabré cluster at the National Laboratory of Advanced Computing (CNCA) and the Costa Rican National High Technology Center (CeNAT).

6. REFERENCES

[1] T. M. Deserno, "Fundamentals of biomedical image processing," in *Biomedical Image Processing*, pp. 1–51, Springer, 2010.

[2] A. Sáenz, S. Calderón, J. Castro, R. Mora, and F. Siles, "Deceived bilateral filter for improving the automatic cell segmentation and tracking in the nf-kb pathway without nuclear staining," in *VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 October 2014*, pp. 345–348, Springer, 2015.

[3] S. Lo, S. Lou, J. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, 1995.

[4] S. Calderón, A. Sáenz, R. Mora, F. Siles, I. Orozco, and M. Buemi, "Dewaff: A novel image abstraction approach to improve the performance of a cell tracking system," in *Bioinspired Intelligence (IWOBI), 2015 4th International Work Conference on*, pp. 81–88, IEEE, 2015.

[5] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in x-ray images," *Medical image analysis*, vol. 36, pp. 41–51, 2017.

[6] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, G. Choy, and S. Do, "Fully automated deep learning system for bone age assessment," *Journal of Digital Imaging*, pp. 1–15, 2017.

[7] A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, and H. Huang, "Bone age assessment of children using a digital hand atlas," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4, pp. 322–331, 2007.

[8] T. S. Sharmila, K. Ramar, and T. S. R. Raja, "Impact of applying pre-processing techniques for improving classification accuracy," *Signal, Image and Video Processing*, vol. 8, no. 1, pp. 149–157, 2014.

[9] T. Nazare, G. P. da Costa, W. Contato, and M. Ponti, "Deep convolutional neural networks and noisy images," in *Iberoamerican Conference on Pattern Recognition (CIARP)*, 2017.

[10] G. B. P. da Costa, W. A. Contato, T. S. Nazare, J. Neto, and M. Ponti, "An empirical study on the effects of different types of noise in image classification tasks," *arXiv preprint arXiv:1609.02781*, 2016.

[11] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*, pp. 1–6, IEEE, 2016.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[13] A. G. Howard, "Some improvements on deep convolutional neural network based image classification," *arXiv preprint arXiv:1312.5402*, 2013.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[16] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting.," *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[17] S. Calderón and M. Zumbado, "Dnlm-iifft: An implementation of the deceived non local means filter using integral images and the fast fourier transform for a reduced computational cost,"

[18] S. Calderón and F. Siles, "Deceived bilateral filter for improving the classification of football players from tv broadcast," in *IEEE 3rd International Conference and Workshop on Bioinspired Intelligence*, 2014.

[19] S. Calderón, D. Moya, J. C. Cruz, and J. M. Valverde, "A first glance on the enhancement of digital cell activity videos from glioblastoma cells with nuclear staining," in *Central American and Panama Convention (CONCAPAN XXXVI), 2016 IEEE 36th*, pp. 1–6, IEEE, 2016.

[20] A. Buades, B. Coll, and J. M. Morel, "Neighborhood filters and pdes," *Numerische Mathematik*, vol. 105, no. 1, pp. 1–34, 2006.

[21] Ž. Emeršič, D. Štepec, V. Štruc, and P. Peer, "Training convolutional neural networks with limited training data for ear recognition in the wild," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 987–994, 2017.

[22] M. Anderson and P. Whitcomb, *DoE Simplified*. Boca Raton, FL, USA: CRC Press, Taylor and Francis Group, 2007.

[23] F. Fallas-Moya, "Object tracking based on hierarchical temporal memory classification," Master's thesis, School of Computer Science - Costa Rican Institute of Technology, Costa Rica, 2015.