58

# Neural Boneprint: Person Identification from Bones using Generative Contrastive Deep Learning

Anonymous Authors

#### ABSTRACT

Forensic person identification is of paramount importance in accidents and criminal investigations. Existing methods based on soft tissue or DNA can be unavailable if the body is badly decomposed, white-ossified, or charred. However, bones last a long time. This raises a natural question: can we learn to identify a person using bone data? We present a novel feature of bones called Neural Boneprint for personal identification. In particular, we exploit the thoracic skeletal data including chest radiographs (CXRs) and computed tomography (CT) images enhanced by the volume rendering technique (VRT) as an example to explore the availability of the neural boneprint. We then represent the neural boneprint as a joint latent embedding of VRT images and CXRs through a bidirectional cross-modality translation and contrastive learning. Preliminary experimental results on real skeletal data demonstrate the effectiveness of the Neural Boneprint for identification. We hope that this approach will provide a promising alternative for challenging forensic cases where conventional methods are limited. The code will be available at \*\*\*.

# CCS CONCEPTS

• Computing methodologies  $\rightarrow$  Computer vision tasks; • Applied computing  $\rightarrow$  Evidence collection, storage and analysis; Life and medical sciences.

# **KEYWORDS**

Person Identification, Skeletal Data, VRT, CXR, Cross Modality, Deep Learning

# 1 INTRODUCTION

"I am not leaving until these bones lead me to wherever my husband is!"

– "Bones", Season 11, Episode 1, FOX 2015

The artwork once yearned there was a technique that could find some subtle clues by analyzing the bones of the victim.

Person identification is the primary and initial concern in accidents and criminal investigations. For a long time, researchers have explored various biological evidence for person identification and authentication, including genetic material deoxyribonucleic acid

Unpublished working draft. Not for distribution.  $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ 

51 52 53 54



Figure 1: When the corpse was deeply skeletonized (highly decomposed, charred, or deliberately destroyed), biological evidence involving soft tissue or DNA is not available. We present Neural Boneprints for person identification by taking the thoracic skeleton as an example.

(DNA) [\[54\]](#page-9-0), faces [\[9,](#page-8-0) [14,](#page-8-1) [45,](#page-8-2) [57\]](#page-9-1), fingerprints [\[36,](#page-8-3) [49\]](#page-9-2), palmprints [\[28,](#page-8-4) [47,](#page-9-3) [61\]](#page-9-4), irises [\[11,](#page-8-5) [49,](#page-9-2) [62\]](#page-9-5), etc.

However, in forensics, when a body/corpse is extensively decomposed, intentionally damaged, or burned, soft tissue markers such as facial features and fingerprints are often neither viable nor extractable for identification. In addition, DNA extraction from such remains is unfortunately extremely challenging due to the progressive degradation of DNA over time [\[17\]](#page-8-6). Beyond the technical difficulties, the identification process is constrained by financial costs, time requirements, and the extent of available DNA databases. The efficacy of DNA analysis heavily relies on the presence of the individual's DNA sequence, or that of close relatives, within these databases. Without prior DNA sequencing and storage, identification becomes significantly hindered, illustrating the limitations of current forensic methodologies in certain scenarios. The intractability and accessibility of biological evidence (e.g. DNA) to skeletons hinders modern forensic identification for skeletons.

Fortunately, there is an important but seldom-exploited basic fact that bones generally persist for a very long time, either in the corpse or in the skeleton. Forensic studies have shown that manual identification of persons through comparison of some skeletal imaging materials between antemortem and postmortem is practical [\[2,](#page-8-7) [13,](#page-8-8) [18,](#page-8-9) [21,](#page-8-10) [42\]](#page-8-11). However, these analyses generally depend on the experience of forensic experts. This raises a natural question: can we learn to identify a person directly from skeletal data? To do this, two fundamental questions need to be answered. First, what skeletal data should we use; second, what kind of features should we learn for identification, and how?

In this paper, we demonstrate that thoracic skeletal data, specifically CXRs and VRT images, are useful for learning to identify directly. We then introduce the Neural Boneprint (NBP), a joint latent embedding extracted through a bidirectional cross-modality translation and contrastive deep learning on CXRs and VRT images, complementing traditional biological metrics. Experimental results

<sup>55</sup>

<sup>56</sup>

<sup>57</sup>

ACM MM, 2024, Melbourne, Australia Anonymous Authors

120 121



Figure 2: The VRT images (blue point) are distributed sparsely  $(P_{VRT}, grav, dashed line)$  while the CXRs (orange point) are easily collected with a dense distribution ( $P_{\text{CXR}}$ , orange, solid line). We design a generative contrastive deep learning approach to learning the identity-based non-linear mapping (orange to blue gradient bidirectional arrows) of different modalities. We realize the identifiable complementary interpolation from one modality to another to obtain the unique feature, which is biometric-like but data-driven and named neural boneprint. VRT2NBP (blue to green gradient trapezoid) obtains NBPs from VRT images.

on real data demonstrate the effectiveness of NBP for person identification in forensics. In summary, the contributions of this work are as follows:

- (i) We present a novel perspective on person identification: learning identifiers called Neural Boneprint (NBP) from skeletal image data, demonstrating the potential of NBP as a biometric-like identifier that complements traditional forensic methods.
- (ii) We present a novel deep learning framework for extracting NBP from CXR and VRT images of the thoracic skeleton. NBP is architecture agnostic, allowing its extraction using different networks.
- (iii) Experimental results on real clinical data demonstrate the effectiveness of NBP in identifying people, achieving a Rank-50 identification accuracy of 84.79% - twice the performance of related state-of-the-art methods.

# 2 MOTIVATION

In this paper, we aim to explore the important but never exploited basic fact that bones generally persist for a very long time, either in the corpse or in the skeleton. We assume that there is an implicit feature boneprint in bones, similar to palmprints and fingerprints, which encodes identity information and is widely present in skeletons and skeletal data.

### 2.1 Which Bones? Thoracic Skeleton

It is worth noting that not each bone on the skeleton can be a suitable candidate for large-scale person identification. Those available bones, such as vertebrae and skulls [\[29,](#page-8-12) [42\]](#page-8-11), contain identifiable boundaries and rich morphology that varies from person to person. In brief, an efficient skeleton-based person identification requires the skeletal data (i) to contain identifiable boundaries and distinct morphology, and (ii) to be easily collected and organized in a largescale matching pool.

While it is commonly acknowledged that human faces and skulls contain identifiable features [\[29\]](#page-8-12), we seek to investigate overlooked

aspects of the skeletal structure that may also hold valuable identity information, and the thoracic skeleton serves as an apt example for this purpose. Thoracic skeletons, including ribs, vertebrae, and sternums [\[19,](#page-8-13) [25,](#page-8-14) [39,](#page-8-15) [56\]](#page-9-6), have been used as manual comparison materials to estimate sex and age due to complex morphology and distinct visual individual differences. Therefore, in this paper, we will explore the thoracic skeleton data to learn thoracic boneprints.

# 2.2 Which Thoracic Skeletal Data? VRT & CXR

Forensic pathologists are often required to preprocess computed tomography (CT) images by volume rendering technology (VRT) [\[16,](#page-8-16) [46\]](#page-9-7) to obtain VRT images for analytical studies. For the corpse, chest VRT can obtain a clear thoracic skeleton to avoid the effects of soft tissue and organ decomposition on skeletal observation. However, it is impractical to construct an adequate pool from VRT images. That is, not everyone has had a CT scan because the risk of radiation-induced cancer is increased by the use of CT [\[4,](#page-8-17) [34\]](#page-8-18), resulting in a paucity of VRT images. As a result, it is unlikely that there is a pre-stored CT image to compare with a query VRT image of an unnamed corpse.

Fortunately, the chest X-ray (CXR) is a regular and inexpensive part of the physical examination. In industrialized countries, there is an average acquisition of 238 erect-view CXRs per 1000 of the population annually [\[5,](#page-8-19) [35\]](#page-8-20). In 2006, approximately 129 million CXRs were obtained in the United States alone [\[33\]](#page-8-21). The CXR is often the first imaging study acquired and remains central to screening, diagnosis, and management of a broad range of conditions [\[40,](#page-8-22) [50\]](#page-9-8). After the COVID-19 pandemic, that number will only be higher. More importantly, CXR images usually contain all the skeletal elements of the chest and their identities. This shows that CXRs can naturally form a huge pool for the identification of thoracic skeletons. Simply using CXRs or synthetic CXRs of corpses may be proposed. However, for corpses that are not fully decomposed, the decomposition and expansion of organs and soft tissues can significantly impact observation and comparison. Furthermore, for skeletonized corpses, the simulated soft tissues and organs in synthetic CXRs can be misleading and thus unsuitable for analysis. The

Neural Boneprint: Person Identification from Bones using Generative Contrastive Deep Learning Metal and Metal ACM MM, 2024, Melbourne, Australia



Figure 3: Top: Query phase. For a query skeletal corpse, we obtain the query NBP from the query skeletal data (VRT image) in the VRT2NBP process. Then match the query NBP with the nearest one in the NBP bank. Bottom: Training phase of the cross-modality translation and cross-modality fusion modules, and construction of the NBP bank via the CXR2NBP process.

question then becomes, how can we exploit the vast amount of CXR data and introduce its rich identification information into the small VRT data? If we can learn boneprint from CXR and VRT images and how?

## 2.3 Comparison Visualization

<span id="page-2-0"></span>

#### Figure 4: VRT images and CXRs from two individuals who possess similar overall thoracic skeletal morphology.

We have visualized VRT images and CXRs from two individuals who possess similar overall thoracic skeletal morphology in Figure [4.](#page-2-0) Although this similarity might obscure obvious differences at first glance, upon closer inspection, subtle variations in chest contours and rib boundary details become apparent. It is noteworthy that even within skeletons exhibiting overall similarity, these minute distinctions are present. In thoracic skeletons displaying significant morphological differences, the disparities would undoubtedly be more prominent.

Challenges. To answer the above questions, several challenges related to the characteristics of skeletal data (VRT, CXR) need to be addressed. The main challenges can be categorized as follows:

- (1) Data Imbalance and 2 Shots. There is also a notable imbalance in the availability of image types. VRT images are relatively scarce and sparse in distribution, in contrast to CXRs, which are abundant and densely distributed. This disparity presents a significant challenge in data handling and analysis. Besides, we have only 2 shots in one category (a single CXR-VRT image pair per individual) in our dataset. This limitation exacerbates the difficulty in extracting reliable boneprint features.
- (2) Large Intra-Class Variation. The considerable modality gap between VRT images and CXRs results in significant intraclass variation. CXRs, in particular, exhibit complex overlaps of various anatomical structures, unlike VRT images that predominantly display skeletal features. Additionally, the varied postures during imaging lead to bone deformation, further complicating the analysis.
- (3) Small Inter-Class Difference. The subtlety of thoracic skeletal differences between individuals poses a formidable challenge. Unlike other identification tasks where inter-class features are distinct, the skeletal features in our case are less conspicuous, sometimes barely discernible to the human eye. This issue is compounded by the physiological similarities in rib numbers and orientations shared among humans, making differentiation based on these features challenging.

Overcoming the above challenges requires sophisticated approaches to image processing, data analysis, and feature extraction to effectively learn and use boneprints for personal identification.

In the next section, we will present a novel deep learning framework to solve these challenges and learn neural boneprints for indentification.

### 3 NEURAL BONEPRINT (NBP)

We aim to use deep learning to learn the boneprints of thoracic skeletons, i.e. neural boneprint (NBP), from VRT images and CXRs for person identification. In particular, to overcome the aforementioned challenges, we propose three modules to learn NBP: (1) Cross-modality Translation. Images are translated into each other's modality to bridge the modality gap and enhance data completeness. (2) Cross-modality Fusion. A dual reconstruction network with contrastive learning fuses fine-grained representations and optimizes inter- and intra-class distances to extract NBPs. (3) NBP-Bank construction. An NBP Bank is constructed from CXR data. Query VRT images are matched against NBPs in the bank for identification.

In this paper, let  $x$  and  $y$  be the VRT image and CXR image,  $T_{x\mu}: x \to y$  and  $T_{\mu x}: y \to x$  are two domain translators. The algorithm ultimately outputs a neural boneprint (NBP) feature through a fusion network F, represented as  $F = NBP(x, y)$ . In the below sections, we will provide a comprehensive description of our NBP algorithm, elucidating its workflow, proposed modules, and design details. The training detail, including the loss function, will also be discussed.

#### 3.1 Cross-Modality Translation

The primary obstacle in converting between VRT images and CXRs is the significant differences in anatomical structures and the distortion of the thoracic skeleton [\[20\]](#page-8-23). This discrepancy creates a significant gap between the two modalities that cannot be effectively bridged by straightforward unidirectional translation [\[60\]](#page-9-9) methods like [\[27\]](#page-8-24).

382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 From a probabilistic standpoint, focusing solely on  $p(CXR|VRT)$ or  $p(Var|CXR)$  in a unidirectional approach hinders our ability to explore the intricate semantic relationships between VRT images and CXRs. To address this, we propose a bidirectional process, that is, simultaneously learning from VRT to CXRs and vice versa. By deliberately aligning their intermediate latent variables in terms of distribution, we establish a preliminary joint latent space between VRT and CXRs, denoted as  $p(VRT, CXR)$ . We argue that this shared latent space enhances the model's capacity to grasp the complex semantic interplay between VRT and CXRs during the cross-modal transformation, as shown in Figure [5.](#page-3-0) Specifically, we employ two transformation networks to model the processes of  $p(CXR|VRT)$ and  $p(Var|CXR)$  separately. We consider the latent variables in the networks' middle layer to follow a multivariate Gaussian distribution and compute the 2-Wasserstein distance  $W$  to measure the divergence between them:

$$
W_2^2(\mathbf{z}_{\text{VRT}}, \mathbf{z}_{\text{CXR}}) = ||\boldsymbol{\mu}_{\text{VRT}} - \boldsymbol{\mu}_{\text{CXR}}||^2
$$
  
+ 
$$
\mathbf{Tr}(\Sigma_{\text{VRT}} + \Sigma_{\text{CXR}} - 2(\Sigma_{\text{VRT}}^{1/2} \Sigma_{\text{CXR}} \Sigma_{\text{VRT}}^{1/2})^{1/2}),
$$
(1)

403 404 405 406 where latent variables are denoted as  $z_{VRT} \sim \mathcal{N}(\mu_{VRT}, \Sigma_{VRT})$  and  $z_{\text{CXR}} \sim \mathcal{N}(\mu_{\text{CXR}}, \Sigma_{\text{CXR}})$ . Two transformation networks are  $T_{xy}$ :  $x \rightarrow y$  and  $\mathsf{T}_{yx} : y \rightarrow x$ .

<span id="page-3-0"></span>

Figure 5: Model the processes of  $p(CXR|VRT)$  and  $p(VRT|CXR)$ to obtain the shared latent space.

In addition, we adopt an extra  $\ell_2$  loss penalty to maintain finegrained personally identifiable information. That is, the real VRT image is paired by identity with the CXR image at the semantic level, so the translated VRT image should be the same as the real VRT image due to the individual identity. In this manner, translated images can be viewed as sampling in the vicinity of the true image in manifold space [\[58\]](#page-9-10).

#### 3.2 Cross-Modality Fusion

After the cross-modality translation, we obtain a real and translated image pair for each individual. Due to the identifiable design, the translated distribution is already close enough to the real distribution in the identity manifold, so we also treat them as real.

We employ a dual input reconstruction network based on contrastive learning to fuse distinguishable skeletal representations to a latent embedding F named neural boneprint. It maps each VRT-CXR image pair to a joint latent embedding in manifold space where the identity information is the primary constraint [\[14,](#page-8-1) [31\]](#page-8-25). It contains a VRT encoder-decoder module, a CXR encoder-decoder module, and a latent fusion module. Each encoder-decoder module is utilized to reconstruct real or translated images for learning the latent fine-grained skeletal representations and then fusing them as neural boneprints. We compute the mean squared error (MSE) between the reconstructed and original images in the pixel space to assist in learning the better fine-grained identical latent features. We also compute the contrastive loss [\[3,](#page-8-26) [10,](#page-8-27) [22\]](#page-8-28) based on the fused embeddings of the real, translated, and the other's real image pairs for minimizing the intra-class distance and maximizing the inter-class distance. During the training phase, we jointly train encoder-decoder modules and the latent fusion module. In the application phase, we ignore decoders and only employ encoders and the latent fusion module with weights frozen.

# 3.3 NBP-Bank Constrtuction

The main idea is that CXRs are widely available and identity-rich, while VRT images are common for skeletal remains. This process aims to create an identification function  $ID(x)$  for efficient identification using a CXR database. Given a CXRs database, we process as following 3 steps:

- (1) Modality Unification: All CXRs images  $\{y_i\}$  are first translated into VRT images  $\{T_{ux}(y_i)\}$  using a translator  $T_{ux}$ .
- (2) Joint Embedding Generation: CXR-VRT pairs  $\{(\mathsf{T}_{yx}(y), y)\}$ are mapped via F to joint latent embeddings (NBPs) that capture cross-modal relationships, incorporating identity information.

(3) NBP-Bank Creation: The NBP-Bank, denoted as C, is constructed as a searchable table with embeddings as keys and identities as values:

$$
C \triangleq \{ (\text{Key} = \text{F}(T_{yx}(y_i), y_i), \text{Value} = \text{ID}(y_i)) \}
$$
 (2)

#### 3.4 Query

As we presented above, the modality unification bridges the CXR and VRT domains, improving the accuracy of identification. The joint embeddings (NBP) can effectively represent cross-modal relationships and identity information. Finally, the constructed NBP bank will enable identification via matching, i.e. table lookup, facilitating efficient matching of new VRT images with identities in the CXR database. In particular, given a query VRT image  $x$ , we proceed as follows 3 steps:

- (1) Modality Translation: Translate  $x$  into its corresponding CXR representation, denoted as  $T_{xy}(x)$ .
- (2) Joint Embedding Generation: Map the image pair  $(x, T_{xy}(x))$ to a joint latent embedding, i.e., NBP =  $F(x_i, T_{x,y}(x_i))$ .
- (3) Identification via Nearest Neighbor Search: Identify the individual associated with  $x$  by performing a nearest neighbor search in the reference NBP bank C. This involves determining the NBP  $\forall c_i \in C$  with the minimum distance d to c, i.e.

$$
ID(x) = \arg\min_{i} d(c, c_i)
$$
\n(3)

#### 4 EVALUATION

#### 4.1 Datasets Setup

All images used in this work are real clinical data collected from \*\* Hospital. We construct two datasets and the first contains VRT-CXR image pairs collected from 1,315 healthy individuals without pulmonary diseases or thoracic skeletal lesions. We randomly split the first dataset into training and test sets. The training set contains VRT-CXR image pairs from 1,052 individuals, representing 80%. The test set contains the others, a total of 263 VRT-CXR image pairs, accounting for 20%. The second dataset contains CXRs collected from 874 healthy individuals. The code and demos will be available at \*\*\*.

#### 4.2 Metrics

Rank-k. As a classical metric in face identification and person re-identification [\[52,](#page-9-11) [59\]](#page-9-12), Rank- $k$  is suitable for our single-shot person identification. In addition, person identification generally involves ethical or legal issues in forensic science. It is also important to simplify the list of candidate victims [\[1\]](#page-8-29). Therefore, we propose to use the Percentile Ranking Rate as the metric.

Percentile Ranking Rate. The percentile ranking rate aims to measure the ability of the model to hit targets in the top  $p$ percent across databases of different scales. Similar to the Rank- $k$ identification rate, we will compute the similarities between the query and the candidates one by one. Then sort the similarities from largest to smallest. If the correct pair is in the top  $p$  percent, the match is considered successful, otherwise, the match fails. The ratio of successful pairs to all pairs is the percentile ranking rate.

#### 4.3 Qualitative Analysis

The VRT images in the test set of the first dataset are used as queries. We apply our approach to two NBP banks, which are organized from our dataset. One is extracted from the CXRs of the first dataset. To test robustness, we introduce the CXRs in the second dataset as distractors and contribute to the second NBP bank with the test set of the first dataset together. More experiments under large-scale databases are prepared in Supplementary.

Alternative Comparison. Since this is the first comprehensive work in this new task, there is no comparable thoracic skeleton identification network. We employ several classical identification models for comparison. In Table [1,](#page-5-0) we compare the performance of Ours and other alternatives in extracting neural skeletal features. The results of IResNet-18 and IResNet-50 [\[15\]](#page-8-30) show that deeper networks did not lead to significant gains. We also compared with a classical cross-modality person re-identification method [\[38\]](#page-8-31).

Cumulative Match Characteristic. We select the six methods with the best performance from all the experiments and plot the cumulative match characteristic curve [\[52\]](#page-9-11), as shown in Figure [6.](#page-4-0)

<span id="page-4-0"></span>

Figure 6: Cumulative match characteristic curve of Rank for different strategies to learn NBP. We achieved state-ofthe-art results along different  $k$  and obtained at least twice the performance of related approaches.

Ablation Study. We explore the ablation study on the extent of the introduction of CMT in Table [2.](#page-5-1) It shows that only using one modality (whether CXR or VRT) cannot be effective. We also explore the ablation study of the positive sample strategies for contrastive learning in the neural boneprint extraction step, as shown in Table [3.](#page-5-2) The Augmented one uses the augmentation strategy of SimCLR [\[10\]](#page-8-27) without cross-modality translation results introduced. Besides, we explore the ablation study on the choice of the reconstruction loss functions with the CMT and the contrastive loss fixed, as shown in Table [4.](#page-5-3) The suboptimal result means the L1 loss can capture fine-grained details than MSE in some specific samples, but MSE can get a better generalization performance for total samples.

Related work Comparison We compared ours with some studies [\[12,](#page-8-32) [26\]](#page-8-33) based on CT images and CXRs for person identification

<span id="page-5-0"></span>

Table 1: The ultimate performance of Ours and other alternatives on learning neural features of bones for identification. The queries are VRT images and the matching pool is NBP bank constructed from the CXR database. A higher value is better.

<span id="page-5-1"></span>

Table 2: Ablation study on introducing the CMT step (CXR only, VRT only, or both). A higher value is better.

<span id="page-5-2"></span>

Table 3: Ablation study on Strategy of determining the positive samples for contrastive learning in the CMF step. The augmented in the second line represents introducing the augmentation strategy of SimCLR [\[10\]](#page-8-27).

<span id="page-5-3"></span>

Table 4: Ablation study on Loss function for Reconstruction (Cross-modality and Contrastive loss are fixed).

in Table [5.](#page-5-4) Our reimplement did not achieve the performance reported in the original work, to ensure a fair comparison, we directly referenced the origin. It shows that our approach contains a larger searchable bank and obtains better performance.

<span id="page-5-4"></span>

Table 5: Our method was experimented with in a larger searchable bank and demonstrated superior performance. A lower  $p$  with a higher accuracy is better.

### 4.4 Quantitative Analysis

We visualize the top 5 samples and the last 5 samples of the identification results after introducing distractors, as shown in Figure [7.](#page-6-0) Neural Boneprint: Person Identification from Bones using Generative Contrastive Deep Learning Machielth ACM MM, 2024, Melbourne, Australia

<span id="page-6-0"></span>

Figure 7: Visualization of robust identification. The query pair consists of the real VRT image and its translated CXR. The candidate pairs consist of real CXRs with their translated VRT images. The orange rectangle represents the ground truth. The red and blue lines describe the rib boundary and the overall skeletal morphology respectively. The top-5 identification results are similar to the query one while the last are not. More visualization results can be found in Supplementary.

With careful comparisons, it can be found that the top five image pairs are highly similar to the query image pairs while the last 5 have low similarity. What's more, the deformations of thoracic skeletons and complex overlaps between VRT images and CXRs make it extremely hard for humans to identify, but it can be done by the proposed approach.

Some unidirectional [\[60\]](#page-9-9) translation methods, such as Pix2Pix [\[27\]](#page-8-24), may be discussed as alternatives in our cross-modality translation step. With that in mind, we also visualize some VRT-CXR image pairs with their translated results by different methods, as shown in

Figure [8.](#page-7-0) VRT images only contain thoracic skeletons while CXRs also contain other anatomical structures besides them. Hence, translating CXRs into VRT images is easier than from VRT images to CXRs. We first compare the skeletal visualization of translated VRT images. The unidirectional results show misaligned bones while the cycle results contain distinct morphology with clear boundaries. An important factor is the different postures while taking the VRT images and the CXRs, which leads to the thoracic skeletons being structurally deformed and having more semantic-level mapping

<span id="page-7-0"></span>

Figure 8: Cycle translation performs better to solve our task. The translated ones by cycle method are highly similar to the real, both in the overall thoracic skeletal structure and in the boundaries and curvature of the ribs.

rather than strictly pixel-level. Besides, the corresponding translated VRT images contain almost the same identifiable rib details and orientations compared with the real. Then we compare the visual performance of generating CXRs. As real VRT images do not contain those anatomical structures, some translated structures similar to organs and muscles may be created during the translation. Despite that, we can ignore the created structures because we only focus on the thoracic skeletons.

#### 5 RELATED WORK

Identification of human remains through antemortem and postmortem imaging materials [\[2\]](#page-8-7) has been widely used.

Manual Methods. For a long time, forensic odontologists have manually utilized dental radiographs taken of the victim before death and manually compared them to dental data from the remains to assist in the identification [\[43,](#page-8-35) [53\]](#page-9-13). Using head and neck CT and MR imaging to obtain identifying information, including dental findings, to identify individuals is also feasible [\[18\]](#page-8-9). Besides, it has been demonstrated that the identification of human remains using visual comparison could be performed by a forensic pathologist with CT experience through 20 cases with antemortem and postmortem CT images, but there is no consensus on the specific number of concordant traits required to establish an identity [\[2\]](#page-8-7). Study [\[51\]](#page-9-14) investigated antemortem and postmortem radiographs of the claviculae and C3-T4 vertebrae to identify skeletons of missing U.S. soldiers from past military operations. The two-dimensional fusion of postmortem computed tomography and antemortem chest radiography makes human identification possible [\[48\]](#page-9-15). In summary, these methods not only rely on manual analysis by experienced forensic experts, but are also difficult to deploy on a large scale.

Automatic Methods. Some automatic identification methods have been proposed recently. Study [\[26\]](#page-8-33) utilized the CXR before

death database to match with CT scan image after death. It obtained an accuracy of 63% within 27 subjects and an average ranking of 10 (total 27, 37.04%) based on the extraction and matching of two types of features, the Bag of Words (BoW) and the Histogram of Gradients (HOG). Study [\[12\]](#page-8-32) utilized morphological erosion to extract rib boundaries and employed Discrete Fourier Transform (DFT) to extract features, which led to an accuracy of 74.07% within rank 15 of 27 subjects. However, the dataset scale of these works is very limited. The performance is not satisfactory and not convincing enough. Study [\[30\]](#page-8-36) realizes the attention localization and alignment of teeth for person identification by semantic segmentation and creating an atlas with landmarks. However, the annotations of the segmentation and the landmarks require manual creation, which is costly both in time and labor. Besides, this method relies on the spatial arrangement of teeth, while skeletonized corpses usually face loss of teeth. Recently, there has been a burgeoning interest in re-identifying CXRs. Some studies [\[32,](#page-8-37) [37,](#page-8-38) [55\]](#page-9-16) achieve patient re-identification through CXRs for proper archival purposes. Identifying solely through CXRs appears to be more straightforward for our task. However, due to the presence of real internal organs in CXRs, these studies are insufficient to demonstrate the capability for skeletal matching. For those not fully decomposed corpses, the decomposition and expansion of organs can greatly affect comparison. For those skeletonized corpses, the simulated soft tissues can be misleading.

# 6 DISCUSSION

We have shown how machine learning techniques can be used to extract neural boneprint from skeletal information that can be used for identification. We hope this will open up new avenues for identity verification and shed new light on the development and cross-fertilization of different disciplines, including forensics and machine learning. There are many open questions that we and the community need to work on together, and the quality of Neural Boneprint features can be improved from many perspectives. First, as Neural Boneprint features are extracted from VRT and CXR images, the quality of bone detection and imaging is particularly important, and we will investigate more effective imaging methods [\[6–](#page-8-39)[8\]](#page-8-40) to improve image quality. Second, although the NBPs extracted from VRT and CXR images in this paper can already be used for identity verification, a natural idea is to use more data modalities (e.g., MRI and PET, etc.) to extract NBPs that contain more information. Third, the cross-modality translation could also be a diffusion-style [\[23,](#page-8-41) [44\]](#page-8-42) approach. The backbone of cross-modality fusion is also various. The localization of thoracic skeletons in original VRT images and CXRs could be deep learning approaches, such as the YOLO [\[41\]](#page-8-43) family. Ultimately, we must prioritize stringent privacy protection and ethical review. Hospital data frequently contain highly sensitive personal information and health records of patients. Therefore, when collecting and utilizing such data, it is imperative to adhere strictly to applicable laws and regulations, ensuring comprehensive protection of patients' privacy rights and interests. To sum up, we start an initial discussion on the proposed task and look forward to more works on social equity and judicial justice based on artificial intelligence.

Neural Boneprint: Person Identification from Bones using Generative Contrastive Deep Learning Metal and Metal ACM MM, 2024, Melbourne, Australia

#### 929 REFERENCES

- <span id="page-8-29"></span>[1] Bradley J Adams and Kenneth W Aschheim. 2016. Computerized dental comparison: A critical review of dental coding and ranking algorithms used in victim identification. Journal of forensic sciences 61, 1 (2016), 76–86.
- <span id="page-8-7"></span>[2] Zabiullah Ali, Nikki Mourtzinos, Bakr B Ali, and David R Fowler. 2020. A Pilot Study Comparing Postmortem and Antemortem CT for the Identification of Unknowns: Could a Forensic Pathologist Do It? Journal of forensic sciences 65, 2 (2020), 492–499.
- <span id="page-8-26"></span>[3] Philip Bachman, R Devon Hjelm, and William Buchwalter. 2019. Learning representations by maximizing mutual information across views. Advances in neural information processing systems 32 (2019).
- <span id="page-8-17"></span>[4] David J Brenner and Eric J Hall. 2007. Computed tomography—an increasing source of radiation exposure. New England journal of medicine 357, 22 (2007), 2277–2284.
- <span id="page-8-19"></span>[5] Erdi Çallı, Ecem Sogancioglu, Bram van Ginneken, Kicky G van Leeuwen, and Keelin Murphy. 2021. Deep learning for chest X-ray analysis: A survey. Medical Image Analysis 72 (2021), 102125.
- <span id="page-8-39"></span>[6] Dongdong Chen, Mike E. Davies, Matthias J. Ehrhardt, Carola-Bibiane Schönlieb, Ferdia Sherry, and Julián Tachella. 2023. Imaging With Equivariant Deep Learning: From unrolled network design to fully unsupervised learning. IEEE Signal Process. Mag. 40, 1 (2023), 134–147.
- [7] Dongdong Chen, Julián Tachella, and Mike E. Davies. 2021. Equivariant Imaging: learning beyond the range space. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 4379–4388.
- <span id="page-8-40"></span>[8] Dongdong Chen, Julián Tachella, and Mike E Davies. 2022. Robust Equivariant Imaging: a fully unsupervised framework for learning to image from noisy and partial measurements. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- <span id="page-8-0"></span>[9] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. 2018. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13. Springer, 428–438.
- <span id="page-8-27"></span>[10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In International conference on machine learning. PMLR, 1597–1607.
- <span id="page-8-5"></span>[11] Yifeng Chen, Cheng Wu, and Yiming Wang. 2019. T-Center: a novel discriminative feature extraction approach for iris recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 9929–9930.
- <span id="page-8-32"></span>[12] Hanni CHO, Thi Thi ZIN, Norihiro SHINKAWA, and Ryuichi NISHII, 2018. Postmortem human identification using chest x-ray and ct scan images. International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association 23, 2 (2018), 51–57.
- <span id="page-8-8"></span>[13] Summer J Decker and Jonathan M Ford. 2019. Forensic personal identification utilizing part-to-part comparison of CT-derived 3D lumbar models. Forensic science international 294 (2019), 21–26.
- <span id="page-8-1"></span>[14] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 4690–4699.
- <span id="page-8-30"></span>[15] Ionut Cosmin Duta, Li Liu, Fan Zhu, and Ling Shao. 2021. Improved residual networks for image and video recognition. In 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 9415–9422.
- <span id="page-8-16"></span>[16] Lars Christian Ebert, Sabine Franckenberg, Till Sieberth, Wolf Schweitzer, Michael Thali, Jonathan Ford, and Summer Decker. 2021. A review of visualization techniques of post-mortem computed tomography data for forensic death investigations. International journal of legal medicine 135, 5 (2021), 1855– 1867.
- <span id="page-8-6"></span>[17] M Fondevila, C Phillips, N Naverán, M Cerezo, A Rodríguez, R Calvo, LM Fernandez, Á Carracedo, and MV Lareu. 2008. Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples. Forensic Science International: Genetics Supplement Series 1, 1 (2008), 26–28.
- <span id="page-8-9"></span>[18] Hideko Fujimoto. 2021. Dental radiographic identification using ante-mortem CT, cone-beam CT, and MRI head and neck assessments. Forensic Imaging 26 (2021), 200465.
- <span id="page-8-13"></span>[19] Nefeli Garoufi, Andreas Bertsatos, Maria-Eleni Chovalopoulou, and Chiara Villa. 2020. Forensic sex estimation using the vertebrae: an evaluation on two European populations. International journal of legal medicine 134 (2020), 2307-2318.
- <span id="page-8-23"></span>[20] Qingbiao Guan, Yutong Xie, Bing Yang, Jianpeng Zhang, Zhibin Liao, Qi Wu, and Yong Xia. 2023. Unpaired Cross-Modal Interaction Learning for COVID-19 Segmentation on Limited CT Images. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 603–613.
- <span id="page-8-10"></span>[21] Gary M Hatch, Fabrice Dedouit, Angi M Christensen, Michael J Thali, and Thomas D Ruder. 2014. RADid: a pictorial review of radiologic identification using postmortem CT. Journal of Forensic Radiology and Imaging 2, 2 (2014), 52–59.
- <span id="page-8-28"></span>[22] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 9729–9738.
- <span id="page-8-41"></span>[23] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. Advances in neural information processing systems 33 (2020), 6840–6851.
- <span id="page-8-34"></span>[24] Elad Hoffer and Nir Ailon. 2015. Deep metric learning using triplet network. In Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3. Springer, 84–92.
- <span id="page-8-14"></span>[25] Martin Hora and Vladimír Sládek. 2018. Population specificity of sex estimation from vertebrae. Forensic science international 291 (2018), 279–e1.
- <span id="page-8-33"></span>[26] Ryudo Ishigami, Thi Thi Zin, Norihiro Shinkawa, and Ryuichi Nishii. 2017. Human identification using X-Ray image matching. In Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol. 1.
- <span id="page-8-24"></span>[27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-toimage translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1125–1134.
- <span id="page-8-4"></span>[28] Wei Li, Lei Zhang, David Zhang, Guangming Lu, and Jingqi Yan. 2010. Efficient joint 2D and 3D palmprint matching with alignment refinement. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 795–801.
- <span id="page-8-12"></span>[29] Yuan Li, Jian Wang, Weibo Liang, Hui Xue, Zhenan He, Jiancheng Lv, and Lin Zhang. 2022. CR-GAN: Automatic craniofacial reconstruction for personal identification. Pattern Recognition 124 (2022), 108400.
- <span id="page-8-36"></span>[30] Yuan Liang, Weikun Han, Liang Qiu, Chen Wu, Yiting Shao, Kun Wang, and Lei He. 2021. Exploring forensic dental identification with deep learning. Advances in Neural Information Processing Systems 34 (2021), 3244–3258.
- <span id="page-8-25"></span>[31] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. 2017. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 212–220.
- <span id="page-8-37"></span>[32] Matthew S Macpherson, Charles E Hutchinson, Carolyn Horst, Vicky Goh, and Giovanni Montana. 2023. Patient reidentification from chest radiographs: an interpretable deep metric learning approach and its applications. Radiology: Artificial Intelligence 5, 6 (2023), e230019.
- <span id="page-8-21"></span>[33] Fred A Mettler Jr, Mythreyi Bhargavan, Keith Faulkner, Debbie B Gilley, Joel E Gray, Geoffrey S Ibbott, Jill A Lipoti, Mahadevappa Mahesh, John L McCrohan, Michael G Stabin, et al. 2009. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950–2007. Radiology 253, 2 (2009), 520–531.
- <span id="page-8-18"></span>[34] Angeliki Neroladaki, Diomidis Botsikas, Sana Boudabbous, Christoph D Becker, and Xavier Montet. 2013. Computed tomography of the chest with modelbased iterative reconstruction using a radiation exposure similar to chest X-ray examination: preliminary observations. European radiology 23 (2013), 360–366.
- <span id="page-8-20"></span>[35] United Nations Scientific Committee on the Effects of Atomic Radiation et al. 2010. Sources and effects of ionizing radiation, united nations scientific committee on the effects of atomic radiation (UNSCEAR) 2008 report, volume I: Report to the general assembly, with scientific annexes A and B-sources. United Nations.
- <span id="page-8-3"></span>[36] Halil İbrahim Öztürk, Berkay Selbes, and Yusuf Artan. 2022. Minnet: Minutia patch embedding network for automated latent fingerprint recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1627–1635.
- <span id="page-8-38"></span>[37] Kai Packhäuser, Sebastian Gündel, Nicolas Münster, Christopher Syben, Vincent Christlein, and Andreas Maier. 2022. Deep learning-based patient reidentification is able to exploit the biometric nature of medical chest X-ray data. Scientific Reports 12, 1 (2022), 14851.
- <span id="page-8-31"></span>[38] Hyunjong Park, Sanghoon Lee, Junghyup Lee, and Bumsub Ham. 2021. Learning by aligning: Visible-infrared person re-identification using cross-modal correspondences. In Proceedings of the IEEE/CVF international conference on computer vision. 12046–12055.
- <span id="page-8-15"></span>[39] Smadar Peleg, Ruth Pelleg Kallevag, Gali Dar, Nili Steinberg, Youssef Masharawi, and Hila May. 2020. New methods for sex estimation using sternum and rib morphology. International Journal of Legal Medicine 134 (2020), 1519–1530.
- <span id="page-8-22"></span>[40] Suhail Raoof, David Feigin, Arthur Sung, Sabiha Raoof, Lavanya Irugulpati, and Edward C Rosenow III. 2012. Interpretation of plain chest roentgenogram. Chest 141, 2 (2012), 545–558.
- <span id="page-8-43"></span>[41] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. 779–788.
- <span id="page-8-11"></span>[42] Ann H Ross, Alicja K Lanfear, and Ashley B Maxwell. 2016. Establishing standards for side-by-side radiographic comparisons. The American Journal of Forensic Medicine and Pathology 37, 2 (2016), 86–94.
- <span id="page-8-35"></span>[43] P Sainio, SM Syrjänen, and S Komakow. 1990. Positive identification of victims by comparison of ante-mortem and post-mortem dental radiographs. The Journal of forensic odonto-stomatology 8, 1 (1990), 11–16.
- <span id="page-8-42"></span>[44] Hiroshi Sasaki, Chris G Willcocks, and Toby P Breckon. 2021. Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. arXiv preprint arXiv:2104.05358 (2021).
- <span id="page-8-2"></span>[45] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition. 815–823.

 

- <span id="page-9-7"></span>[46] Wolf Schweitzer, Thomas Ruder, Michael Thali, and Helmut Ringl. 2015. Skull fractures in post-mortem CT: VRT, flat and skin surface projections in comparison. Journal of Forensic Radiology and Imaging 3, 4 (2015), 214–220.
- <span id="page-9-3"></span>[47] Huikai Shao, Dexing Zhong, and Xuefeng Du. 2019. Efficient deep palmprint recognition via distilled hashing coding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 0–0.
- <span id="page-9-15"></span>[48] Norihiro Shinkawa, Toshinori Hirai, Ryuichi Nishii, and Nobuhiro Yukawa. 2017. Usefulness of 2D fusion of postmortem CT and antemortem chest radiography studies for human identification. Japanese Journal of Radiology 35 (2017), 303– 309.
- <span id="page-9-2"></span>[49] Shashi Shreya and Kakali Chatterjee. 2024. Latent fingerprint and Iris fusion for enhancement of performance of human identification system. Expert Systems with Applications 235 (2024), 121208.
- <span id="page-9-8"></span>[50] Anouk M Speets, Sandra Kalmijn, Arno W Hoes, Yolanda der Graaf, Hugo M Smeets, and Willem P Th M Mali. 2005. Frequency of chest radiography and abdominal ultrasound in the Netherlands: 1999–2003. European journal of epidemiology 20 (2005), 1031–1036.
- <span id="page-9-14"></span>[51] Carl N Stephan, Allysha P Winburn, Alexander F Christensen, and Andrew J Tyrrell. 2011. Skeletal identification by radiographic comparison: blind tests of a morphoscopic method using antemortem chest radiographs. Journal of forensic sciences 56, 2 (2011), 320–332.
- <span id="page-9-11"></span> [52] Han Sun, Zhiyuan Chen, Shiyang Yan, and Lin Xu. 2019. Mvp matching: A maximum-value perfect matching for mining hard samples, with application to person re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 6737–6747.
- <span id="page-9-13"></span> [53] Niveditha Thampan, R Janani, R Ramya, R Bharanidharan, A Ramesh Kumar, and K Rajkumar. 2018. Antemortem dental records versus individual identification. Journal of forensic dental sciences 10, 3 (2018), 158.
- <span id="page-9-0"></span> [54] Rosemary S Turingan, Jessi Brown, Ludmila Kaplun, Jake Smith, Jenna Watson, Derek A Boyd, Dawnie Wolfe Steadman, and Richard F Selden. 2020. Identification of human remains using Rapid DNA analysis. International Journal of Legal

Medicine 134 (2020), 863–872.

- <span id="page-9-16"></span>[55] Yasuyuki Ueda and Junji Morishita. 2023. Patient identification based on deep metric learning for preventing human errors in follow-up X-ray examinations. Journal of Digital Imaging 36, 5 (2023), 1941–1953.
- <span id="page-9-6"></span> [56] Emeline Verna, Marie-Dominique Piercecchi-Marti, Kathia Chaumoitre, Christophe Bartoli, Georges Leonetti, and Pascal Adalian. 2013. Discrete traits of the sternum and ribs: a useful contribution to identification in forensic anthropology and medicine. Journal of Forensic Sciences 58, 3 (2013), 571-577.
- <span id="page-9-1"></span>[57] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. 2018. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 5265–5274.
- <span id="page-9-10"></span>[58] Tom White. 2016. Sampling generative networks. arXiv preprint arXiv:1609.04468 (2016).
- <span id="page-9-12"></span>[59] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. 2019. Unsupervised person re-identification by soft multilabel learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2148–2157.
- <span id="page-9-9"></span>[60] Jiaojiao Zhang, Shuo Zhang, Xiaoqian Shen, Thomas Lukasiewicz, and Zhenghua Xu. 2024. Multi-ConDoS: Multimodal Contrastive Domain Sharing Generative Adversarial Networks for Self-Supervised Medical Image Segmentation. IEEE Transactions on Medical Imaging 43, 1 (2024), 76–95. [https://doi.org/10.1109/](https://doi.org/10.1109/TMI.2023.3290356) [TMI.2023.3290356](https://doi.org/10.1109/TMI.2023.3290356)
- <span id="page-9-4"></span>[61] Kai Zhao, Lei Shen, Yingyi Zhang, Chuhan Zhou, Tao Wang, Ruixin Zhang, Shouhong Ding, Wei Jia, and Wei Shen. 2022. BézierPalm: A free lunch for palmprint recognition. In European Conference on Computer Vision. Springer,  $19 - 36.$
- <span id="page-9-5"></span>[62] Zijing Zhao and Ajay Kumar. 2017. Towards more accurate iris recognition using deeply learned spatially corresponding features. In Proceedings of the IEEE international conference on computer vision. 3809–3818.