Active Latent Space Shape Model: A Bayesian Treatment of Shape Model Adaptation with an Application to Psoriatic Arthritis Radiographs

Adwaye Rambojun [1] William Tillett [1,2] Tony Shardlow [1] Neill D. F. Campbell [1]

[1] University of Bath [2] Royal National Hospital for Rheumatic Diseases

Abstract

Shape models have been used extensively to regularise segmentation of objects of interest in images, e.g. bones in medical x-ray radiographs, given supervised training examples. However, approaches usually adopt simple linear models that do not capture uncertainty and require extensive annotation effort to label a large number of set template landmarks for training. Conversely, supervised deep learning methods have been used on appearance directly (no explicit shape modelling) but these fail to capture detailed features that are clinically important.

We present a supervised approach that combines both a non-linear generative shape model and a discriminative appearance-based convolutional neural network whilst quantifying uncertainty and relaxes the need for detailed, template based alignment for the training data. Our Bayesian framework couples the uncertainty from both the generator and the discriminator; our main contribution is the marginalisation of an intractable integral through the use of radial basis function approximations. We illustrate this model on the problem of segmenting bones from Psoriatic Arthritis hand radiographs and demonstrate that we can accurately measure the clinically important joint space gap between neighbouring bones.

1. Introduction

Psoriatic Arthritis (PsA) is an inflammatory disease that affects the joints of the hands and feet. Radiographic assessment of joint damage is an essential part of PsA diagnosis and treatment, during which each joint of the hand is manually "scored" (a visual assessment) for damage using standardised techniques [16, 22, 29, 31] (see Figure 1 for an illustration). This task is both time consuming and has a high variance in quality due to the subjective evaluation of clinicians; consequently computer vision methods aimed at speeding this process up have been investigated [10, 11, 18, 28].

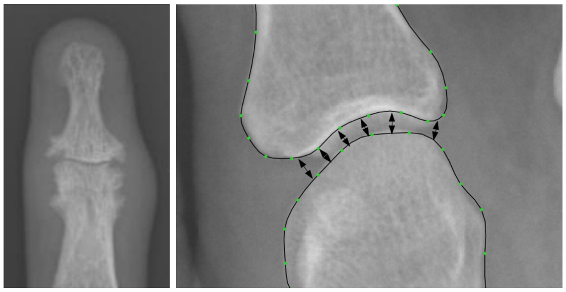

Figure 1. The right Figure (courtesy of Tillett et al. [25]) shows a distal interphalangeal joint suffering from erosion, joint space narrowing (JSN) and osteoproliferation. The figure on the left shows the joint space being measured through the use of bounding curves.

Previous work has considered the use of appearance-based Neural Network models (NNs) to detect damage in x-rays [18, 23]. However, these approaches tend to underperform due to the imbalances in datasets towards healthy examples; they struggle to detect damage that manifests as fine-scaled features. Consequently, they have been observed to be better at detecting extreme examples [6]. It is easier to extract shape information than it is to interpret fine scale texture information. Hence, models that explicitly combine shape and appearance have had more success at detecting geometric shape features than their deep learning counterparts had at detecting texture based features [10, 11, 28]. More recently, clinical studies have been performed to assess methods relying on shape and appearance [13, 21].

We wish to combine appearance-based NNs and Statistical Shape Models (SSMs) to measure joint space, as shown in Figure 1, by segmenting bones in a planar x-ray radiograph. Shape tends to express itself through long range correlations; these are difficult to capture solely by pixel based models that assume a much shorter correlation range. Hence, segmentation methods that rely solely on appearance based NNs are not very good at capturing shape information. Shape models, on the other hand, are better at extracting fine-scale shape information, but still require the use of texture information to ensure accurate fitting. We believe that better performance can be achieved when used alongside deep networks which have been shown to discriminate and model texture well [19, 27, 30].

Capturing the uncertainty in the output of a model is critically important in a medical application (to support downstream clinical decision making) and is one of our primary goals in this paper. We consider the problem of error propagation from both the SSM and the deep learning network. We propose to view the whole problem from a Bayesian viewpoint, where we wish to integrate out model parameters at fitting time; we use a Radial Basis Function (RBF) approximation to tackle the intractable integral arising from such a formulation.

The marginalisation we perform requires neither the shape model nor the deep learning network to themselves be the result of a Bayesian model, but rather assumes that we have some measure of uncertainty in both of their outputs. It is this error that we wish to propagate in the fitting step of our model.

Established statistical models of shape and appearance (e.g. Active Shape Models [4]) assume a generative model as the result of a linear mapping from some learned latent space. This assumption can be relaxed, in which case a lower dimensional latent space is able to represent a richer variety of shapes [8]. We hence use a Gaussian Process Latent Variable Model (GPLVM) [12] as a statistical model of shape.

A second problem we look at is that of data annotation. SSMs usually require careful placement of landmarks on images for training. We propose to treat shape as a continuous curve in 2D. This allows us to align landmarks based on their geometric information. Annotation then becomes an exercise of delineation which can be faster and more robust than landmark placement; this saves time and cost for annotation from trained clinical experts.

We provide related work and background information in section 2, where we introduce both the way shape models are fitted to images, and also GPLVMs and how they can be used as shape priors. Section 3 presents our model, which we refer to as a Active Latent Space Shape Model (ALSSM). We perform the marginalisation of model parameters in section 3.2 and provide error estimates in section 3.3. We then show experimental results in section 4; these consist of a 10-fold validation exercise where we analyse the effect of the appearance discriminator and the shape model on the accuracy of the model output when compared to ground truth.

2. Related Work and Background

Statistical Shape Model Fitting: We consider a continuous image intensity function $u : \mathbb{R}^2 \to [0, 255]$ having edge set Γ. The edge set is defined as the jump set of u and manifests itself as areas of high gradients in u. The boundaries of objects of interest (i.e. bones) tend to be a subset of Γ.

The SSM of Cootes et al. [4] generates a shape vector $\mathbf{F} = (x_0, y_0, ..., x_{D-1}, y_{D-1})$ representing a discretised curve with discretisation number D bounding an object of interest. In the literature, this vector can be augmented to include texture information $\boldsymbol{u}_d$ around the spatial points (x_d, y_d), $d \in [0, D)$, in which case, these are referred to as Active Appearance Models (AAMs) and $\mathbf{F} = (x_0, y_0, \boldsymbol{u}_0, ..., x_{D-1}, y_{D-1}, \boldsymbol{u}_{D-1})$. Without loss of generality, we assume $\mathbf{F} \in \mathbb{R}^P$, where $P = 2D$ in the case of SSMs. The linear shape generation process is learned from training data $\mathbf{Z} \in \mathbb{R}^{N \times P}$ consisting of N discretised curve outlines and is given by

$$\mathbf{F} = \bar{\boldsymbol{z}} + \sum_{q=0}^{Q-1} t_q \boldsymbol{v}_q \ , \tag{1}$$

where $\bar{\boldsymbol{z}}$ is the mean feature vector and $\boldsymbol{v}_q$ are the Q orthogonal variation modes or eigenvectors of $\frac{1}{N}(\mathbf{Z} - \bar{\mathbf{Z}})^t(\mathbf{Z} - \bar{\mathbf{Z}})$. $\boldsymbol{t} = (t_0, ..., t_{q-1})$ can be thought of as a latent variable that gets mapped linearly onto shape or appearance space.

Shape models are trained on data that is made invariant to similarity transforms. The fitting step also involves a similarity transform $\mathcal{T}$. Generally, statistical appearance and shape models are fitted to new images by minimising a cost function of the form

$$\sum_{d=0}^{D-1} V_u\big(\mathcal{T} \circ (x_d, y_d), \boldsymbol{u}_d\big) \ , \tag{2}$$

where $V_u(\cdot)$ is some potential, with respect to the latent space parameter and the pose parameters. When dealing purely with shape models, we consider the edge potential $V_u : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$V_u(x, y) \propto \log\left(\mathbb{P}\{(x, y) \in \Gamma\}\right) \tag{3}$$

that is minimal at locations along the edge set Γ and increases away from Γ.

The model fitting relies on gradient updates that are local in nature. Hence, model parameter initialisation, in particular those relating to pose, becomes very important. This is the main barrier to completely automating shape model fitting. Object positioning can be inferred from global search methods to help with this initialisation in datasets that have a strong prior on object placement in images [3, 7].

Aligned Training Data: The training data for SSMs consist of coordinates or landmarks $\{x_d, y_d\}$, for each of N examples, that need to be in precise alignment. This makes the data annotation process expensive, especially in cases where the number of coordinates D needs to be high in order to capture fine scale features. One way to make this process faster is to instead perform the landmark alignment after the data annotation process; Campbell et al. [2] performs this, for font outlines, through an energy based alignment of training landmarks which considers the geometric shape features.

Uncertainty Quantification: The Probabilistic Appearance Model of Kruger et al. [9] is an example of work that seeks to incorporate uncertainty quantification. The fitting step is interpreted as a Bayesian one and a Maximum a Posteriori (MAP) solution is found for the pose and model hyper parameters. The appearance is treated as a Gaussian with mean given by $\bar{\boldsymbol{z}}+\sum_{q=0}^{Q-1} t_q \boldsymbol{v}_q$. The cost function being minimised while fitting the appearance model to a template is essentially L-2 in nature through the use of Gaussian distributions that model the interaction between the shape and its latent space parameters.

For an image u, our proposed solution to the Bayesian problem of fitting an appearance or shape model with latent parameters $\boldsymbol{t}$ and pose parameters $\mathcal{T}$, is the minimiser of the marginal log-likelihood given by $\log p(u \,|\, \boldsymbol{t}, \mathcal{T})$; we seek the set of appearance or shape parameters $\boldsymbol{t}$ and pose parameters $\mathcal{T}$ that best explain the observed image.

The marginal log-likelihood is an expectation that constitutes an integral made intractable by the edge potential. In the literature, this is usually addressed by approximating the posterior distribution with some variational distribution that makes the integration tractable. The optimal form of the variational distribution is one that is proportional to the posterior. As far as we are aware, there are no explicit error estimates for this approximation. We believe that this can be solved for our case by instead approximating the intractable integral with a Radial Basis Function (RBF) expansion.

GPLVMs and Shape Modelling: GPLVMs [12] are probabilistic generative models. They assume that data $\boldsymbol{z} \in \mathbb{R}^P$ is generated non-linearly from a latent variable $\boldsymbol{t} \in \mathbb{R}^Q$ existing in some learned latent space through $\boldsymbol{z} = g(\boldsymbol{t}) + \epsilon$. Here, $g(\cdot) \sim \mathcal{GP}$ is a zero mean Gaussian Process with independent and identically distributed outputs having covariance kernel $\kappa_\theta(\cdot,\cdot)$ and ϵ is zero mean Gaussian noise with variance β^{-1}. The latent point parameters $\mathbf{T} = [..., \boldsymbol{t}_n, ...]$ for the N training data points are found by minimising the negative marginal log-likelihood given by

$$L(\mathbf{T}, \theta) = \frac{PN}{2}\log(2\pi) + \frac{P}{2}\log|\mathbf{K}| + \frac{1}{2}\mathrm{tr}(\mathbf{K}^{-1}\mathbf{Z}\mathbf{Z}^t) \quad (4)$$

where $\mathbf{K}_{i,j} = \kappa(\boldsymbol{t}_i, \boldsymbol{t}_j) + \frac{1}{\beta}\mathbb{I}_i^j$ and θ are the kernel hyperparameters. GPLVMs have been shown to perform well as generative models of parametric curve representations of shape [2, 15]. Priscariu et al. [15] use the GPLVM posterior mean

$$\boldsymbol{\mu}_{\text{post}}(\boldsymbol{t}) = \kappa(\mathbf{T}, \boldsymbol{t})^t \mathbf{K}^{-1}\mathbf{Z} \in \mathbb{R}^P. \quad (5)$$

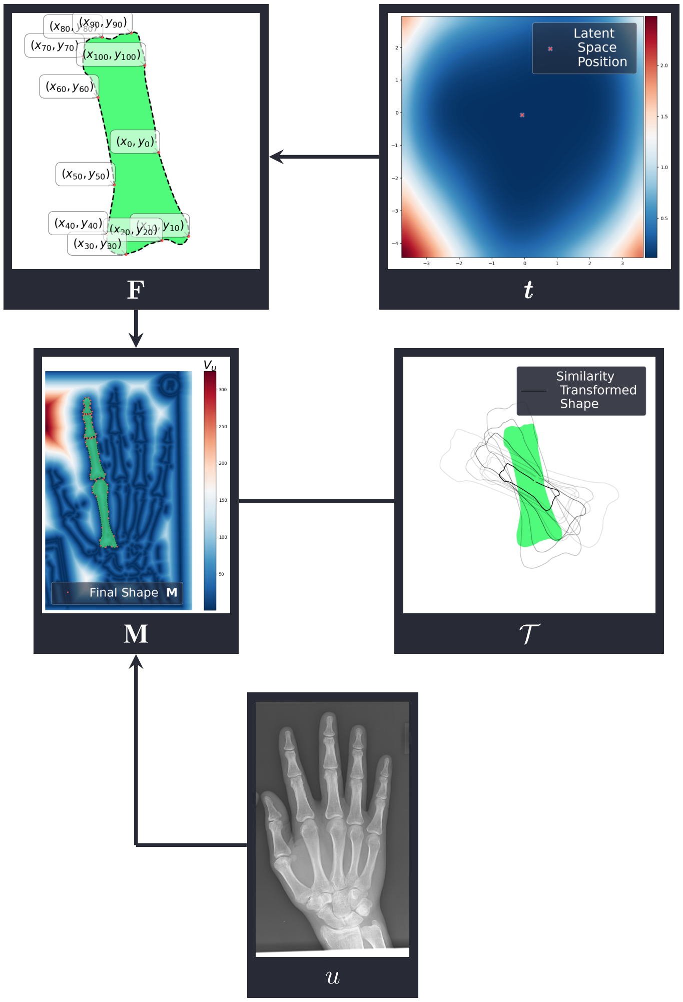

Figure 2. The generative ALSSM. The latent space parameters $\boldsymbol{t}$ generate a shape **F**; this is transformed by $\mathcal{T}$ using the pose parameters to get a curve **M** that interacts with the edge potential $V_u(\cdot)$. The arrows shows the interdependencies of each variable. We perform the marginalisation of **F** and **M** in section 3.2.

to regularise a level set segmentation task. Di Martino et al. [5] investigates the use of GPLVMs as generative models of shape when operating on the pixel domain and show that better performance can be achieved when GPLVMs are paired with Deep Belief Networks.

3. ALSSM

3.1. Model Set-up

The Active Latent Space Shape Model consists of a shape generation process from the latent space variables $\boldsymbol{t}$. The generated shape **F** is transformed via a similarity map $\mathcal{T}$ and matched to the shape $\mathbf{M} = (x_0^{\mathrm{M}}, y_0^{\mathrm{M}}, ..., x_{D-1}^{\mathrm{M}}, y_{D-1}^{\mathrm{M}})$ present in an image u. These are related to each other via the graphical model in Figure 2 that constitutes the following conditional distributions.

Prior on shape: We use a GPLVM to model the interaction between the shape and the latent space. Thus, the shape prior $p(\mathbf{F} \,|\, \boldsymbol{t})$ is the predictive posterior from the trained GPLVM; it takes the form of a Gaussian with mean $\boldsymbol{\mu}(\boldsymbol{t}) = (x^t_0, y^t_0, .., x^t_{D-1}, y^t_{D-1})$ and covariance matrix variance $\sigma^2(\boldsymbol{t})\, \mathbf{I}_Q$. For a GPLVM trained on data $\mathbf{Z}$ with latent space positions $\mathbf{T}$, these are

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{t}) &= \kappa(\mathbf{T}, \boldsymbol{t})^t (\mathbf{K} + \beta^{-1}\mathbf{I}_N)^{-1}\mathbf{Z} \in \mathbb{R}^P, \text{ and} \\ \sigma^2(\boldsymbol{t}) &= \kappa(\boldsymbol{t}, \boldsymbol{t}) - \kappa(\mathbf{T}, \boldsymbol{t})^t (\mathbf{K} + \beta^{-1}\mathbf{I}_N)^{-1} \kappa(\mathbf{T}, \boldsymbol{t}).\end{aligned} \tag{6}$$

Data Likelihood: For ease of notation, we define

$$f_u(x, y) := \exp\left(-V_u(x, y)\right) \ . \tag{7}$$

We set $p(u \,|\, \mathbf{M}) = \prod_{d=0}^{D-1} p(u \,|\, x^{\mathrm{M}}_d, y^{\mathrm{M}}_d)$. We use a discriminative texture model

$$p(x^{\mathrm{M}}_d, y^{\mathrm{M}}_d \,|\, u) = f_u(x^{\mathrm{M}}_d, y^{\mathrm{M}}_d) \tag{8}$$

that computes the probability of the coordinates $(x^{\mathrm{M}}_d, y^{\mathrm{M}}_d)$ being in the edge set Γ as per equation (3). We can invert this probability using Bayes Theorem. Setting $p(x^{\mathrm{M}}_d, y^{\mathrm{M}}_d)$ to be the uniform distribution on the image domain, we have

$$p(u \mid \mathbf{M}) \propto \exp\left(-\sum_{d=0}^{D-1} V_u(x^{\mathrm{M}}_d, y^{\mathrm{M}}_d)\right) . \tag{9}$$

We define $V_u(x, y)$ to be the minimum distance between (x, y) and the image edge set Γ. To obtain the edge set, we use a U-net region discriminator as described in section 4.3.

Shape Matching Term: This compares the curve appearing in the image to a transformed version of the generated shape using a Gaussian given by

$$p(\mathbf{M} \,|\, \mathbf{F}, \mathcal{T}) = \prod_{d=0}^{D-1} \mathcal{N}\big((x^{\mathrm{M}}_d, y^{\mathrm{M}}_d) \mid \mathcal{T} \circ (x_d, y_d), \gamma^2 \mathbf{I}_2\big) \, . \tag{10}$$

The term γ^2 captures the error present in trying to discern the shape appearing in the image.

Whole Model: We seek to maximise

$$\begin{aligned} p(u \,|\, \boldsymbol{t}, \mathcal{T}) &= \iint p(\mathbf{M}, \mathbf{F}, u \,|\, \boldsymbol{t}, \mathcal{T})\, \mathrm{d}\mathbf{F}\, \mathrm{d}\mathbf{M} \\ &= \iint p(u \,|\, \mathbf{M})\, p(\mathbf{M} \,|\, \mathbf{F}, \mathcal{T})\, p(\mathrm{F} \,|\, \boldsymbol{t})\, \mathrm{d}\mathbf{F}\, \mathrm{d}\mathbf{M} \ , \end{aligned} \tag{11}$$

where we assume that the shape parameters are independent to the pose parameters. This is similar to the minimisation performed by Kruger et al. [9].

3.2. Bayesian Marginalisation

Marginalising $\mathbf{F}$ and $\mathbf{M}$ in the model in Figure 2 allows us to capture the errors from each distribution. In this section, we show that by approximating $p(u \,|\, \mathbf{M})$ with a linear combination of Gaussians, we find that the final cost function is $p(u \,|\, \mathbf{M})$ convolved by a Gaussian and evaluated at the transformed coordinates $\mathcal{T} \circ (x^t_d, y^t_d)$. We denote the convolution of a function f with a Gaussian with mean zero and variance σ^2 as $f \star \mathcal{N}_{\sigma^2}$.

With the above distributions for $p(\mathbf{M} \,|\, \mathbf{F}, \mathcal{T})$ and $p(\mathbf{F} \,|\, \mathcal{T})$, and denoting $\mathrm{F}_d = (x_d, y_d)$, $\mathrm{M}_d = (x^{\mathrm{M}}_d, y^{\mathrm{M}}_d)$ and $\mu_d(\boldsymbol{t}) = (x^t_d, y^t_d)$, equation (11) becomes

$$\begin{aligned} & p(u \,|\, \boldsymbol{t}, \mathcal{T}) \\ & \propto \prod_{d=0}^{D-1} \int f_u(\mathrm{M}_d)\, p(\mathrm{M}_d \,|\, \mathrm{F}_d, \mathcal{T})\, p(\mathrm{F}_d \,|\, \boldsymbol{t})\, \mathrm{dF}_d\, \mathrm{dM}_d \\ & \propto \prod_{d=0}^{D-1} \int f_u(\mathrm{M}_d)\, \mathcal{N}\big(\mathrm{M}_d \,|\, \mathcal{T} \circ \mu_d(\boldsymbol{t}), \big(\gamma^2 + \sigma^2(\boldsymbol{t})\big)\, \mathbf{I}_2\big)\, \mathrm{dM}_d \end{aligned} \tag{12}$$

where $p(u \,|\, \mathrm{M}_d)$ makes the above integral intractable.

By definition, the values of $f_u(\mathrm{M}_d)$, on the image lattice $\{\mathcal{X}_h \in \mathbb{R}^2 : h = 0, ..., H-1\}$, are known. Here the image lattice are the uniformly spaced locations at which the image intensity is sampled. We can therefore approximate it using an RBF interpolant of the form

$$\mathcal{I}(\mathrm{M}_d) = \sum_{h=0}^{H-1} w_h\, \psi_h(\mathrm{M}_d) \tag{13}$$

where $\psi_h(\mathrm{M}_d) = \mathcal{N}\big(\mathrm{M}_d \,|\, \mathcal{X}_h, v^2\mathbf{I}_2\big)$ and w_h are scalar weights. This approximation of $f_u(\mathrm{M}_d)$ by a Gaussian allows us to marginalise M_d; we perform the same marginalisation as in equation (12), but now with the interpolant in lieu of $f_u(\mathrm{M}_d)$:

$$\begin{aligned} & \int \mathcal{I}(\mathrm{M}_d)\, \mathcal{N}\big(\mathrm{M}_d \,\big|\, \mathcal{T} \circ \mu_d(\boldsymbol{t}), \gamma^2 + \sigma^2(\boldsymbol{t})\big) \mathrm{dM}_d \\ & = \sum_{h=0}^{H-1} w_h \int \psi_h(\mathrm{M}_d)\, \mathcal{N}\big(\mathrm{M}_d \,\big|\, \mathcal{T} \circ \mu_d(\boldsymbol{t}), \gamma^2 + \sigma^2(\boldsymbol{t})\big)\, \mathrm{dM}_d \\ & = \sum_{h=0}^{H-1} \frac{w_h}{\sqrt{2\pi(v^2 + \gamma^2 + \sigma^2(\boldsymbol{t}))}} \exp\left(-\frac{\|\mathcal{T} \circ \mu_d(\boldsymbol{t}) - \mathcal{X}_h\|^2}{2(v^2 + \gamma^2 + \sigma^2(\boldsymbol{t}))}\right) . \end{aligned} \tag{14}$$

Noting that

$$\psi_h \star \mathcal{N}_{\sigma^2(\boldsymbol{t}) + \gamma^2}(\cdot) = \mathcal{N}(\cdot \mid \mathcal{X}_h, (v^2 + \gamma^2 + \sigma^2(\boldsymbol{t}))\mathbf{I}_2) \ , \tag{15}$$

due to the linearity of the convolution operator, equation (14) is the convolution of $\mathcal{I}$ with a zero mean Gaussian with variance $\sigma^2(\boldsymbol{t}) + \gamma^2$, that is

$$\begin{aligned} & \mathcal{I} \star \mathcal{N}_{\sigma^2(\boldsymbol{t}) + \gamma^2}(\mathcal{T} \circ \mu_d(\boldsymbol{t})) \\ & = \int \mathcal{I}(\mathrm{M}_d)\, \mathcal{N}\big(\mathrm{M}_d \,\big|\, \mathcal{T} \circ \mu_d(\boldsymbol{t}), \gamma^2 + \sigma^2(\boldsymbol{t})\big)\, \mathrm{dM}_d \ . \end{aligned} \tag{16}$$

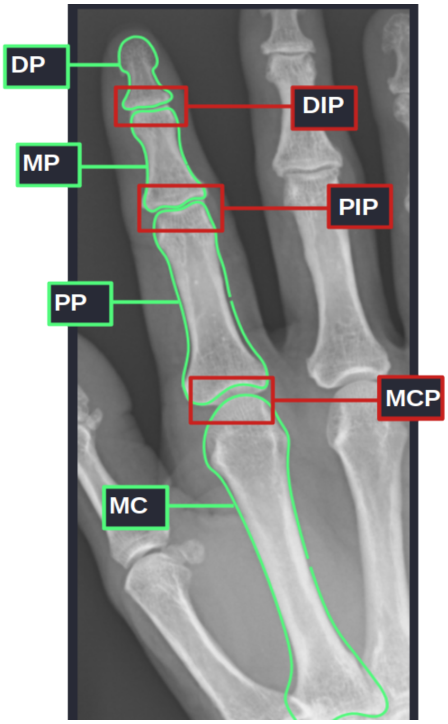

Figure 3. The ALSSM output is given in green. From top to bottom the bones of the right index fingers are the distal phalanx (DP), the middle phalanx (MP), the proximal phalanx (PP) and the metacarpal (MC); and the joints between these are the distal interphalangeal joint (DIP), the proximal interphalangeal joint (PIP), and the metacarpophalangeal joint (MCP).

The final integral can thus be approximated using the result $f_u(\cdot) \star \mathcal{N}_{\sigma^2(\boldsymbol{t})+\gamma^2}(\mathcal{T} \circ \mu_d(\boldsymbol{t}))$.

3.3. Analytical Error Estimates

The error is split into two parts and comes from each RBF expansion. Rambojun et al. [17] show that the error from each part goes to zero as the discretisation H of the image lattice goes to infinity. Hence, this approximation works well for high resolution images. We provide a summary of the argument made by Rambojun et al. [17] in Appendices D and E in the supplemental material.

3.4. Numerical Approximation of the Objective

We recover an approximation of $p(u \,|\, \boldsymbol{t}, \mathcal{T})$ that is given by

$$\hat{p}(u \,|\, \boldsymbol{t}, \mathcal{T}) = f_u \star \mathcal{N}_{\sigma^2(\boldsymbol{t})+\gamma^2}\big(\mathcal{T} \circ \mu_d(\boldsymbol{t})\big) \; . \tag{17}$$

The derivative of the above expression with respect to $\boldsymbol{t}$ cannot be analytically computed. Hence, we replace equation (17) with

$$\tilde{p}(u \,|\, \boldsymbol{t}, \mathcal{T}) = \exp\left(-\frac{\sum_{d=0}^{D-1} V_u\big(\mathcal{T} \circ \mu_d(\boldsymbol{t})\big)}{1 + \sigma^2(\boldsymbol{t}) + \gamma^2}\right) \tag{18}$$

which we maximise instead of $\hat{p}(u \,|\, \boldsymbol{t}, \mathcal{T})$ directly. We provide an empirical justification for this approximation in the supplemental material appendix A. In our experiments, we minimise the negative log of equation (18) which is given by

$$E_0(\boldsymbol{t}, \mathcal{T}) = \sum_{d=0}^{D-1} V_u(\mathcal{T} \circ \mu_d(\boldsymbol{t})) + \sigma^2(\boldsymbol{t}) \; . \tag{19}$$

4. Experiments

In this section, we investigate the effect of the latent space dimension, the edge potential function, and the type of mapping from the latent space to shape space. We describe how we use the GPLVM kernel to switch from a linear and a non-linear mapping in section 4.2. We report the deviation from the ground truth of the bone outline and the joint space for each bone in the right index finger in a 10-fold validation exercise. Figure 3 shows the bones being segmented and the joints being measured. We perform further experiments in in the supplemental material appendix B, where we also investigate the effect of the posterior variance term in equation (19).

To obtain the Joint Space Width (JSW) from two neighbouring curve outlines, we first find the distance of each point in the bottom bone outline from the top bone outline via the distance transform of the top curve. Then we compute the average distance from the top bone outline of 10 successive coordinates in the bottom bone outline. That is, we perform a convolution on the array containing the distance of each coordinate of the bottom curve from the top curve with a kernel of size 10. We take the joint space to be the minimum of these averages.

4.1. Dataset

For real world evaluation, we used $N = 101$ radiographs from an observational research cohort [1]. The radiographs all had Sharp Van der Heijde and Ratingen scores of 0 along the index finger [29, 31]. All patients fulfilled the CASPAR criteria for PsA [24]. The principles of the Declaration of Helsinki were followed and ethical approval was obtained from the National Research Ethics Services (NRES) Committee South West Wales Panel D. All patients included in this study gave full written informed consent for participation.

The outlines of the MP, the PP, the MP and the DP of the right index finger was delineated by an expert Rheumatologist using an in-house annotation software. We treat these points as coming from the discretisation or sampling of a parametric curve $\mathbf{r}$, and as such, they were upsampled to the same discretisation $D = 110$ using a Fourier Curve representation [15]:

$$\mathbf{r}(s) = \begin{pmatrix} \sum_{j=0}^{N_f-1} a_j \cos(2j\pi s) + b_j \sin(2j\pi s) \\ \sum_{j=0}^{N_f-1} c_j \cos(2j\pi s) + d_j \sin(2j\pi s) \end{pmatrix}. \tag{20}$$

Alignment: They were then made invariant to similarity transforms by warping them onto the level set of the distance transform of some template example. The sampled points on these curves were then put in shape correspondence. Given two curves $\mathbf{r}_1, \mathbf{r}_2 : [0,1] \to \mathbb{R}^2$, two sampled points $\mathbf{r}(s_1), \mathbf{r}(s_2)$ would be in shape correspondence if they have the same geometry at s_1 and s_2 respectively. Let $\mathbf{S} \in \mathbb{R}^{N\times D}$ be the sampling point data matrix where $s_{n,d}$ is the d-th sampling location of the n-th data point. We want want the derivatives at coordinate d to be the same across the curve dataset. Campbell et al. [2] relax this requirement, and instead minimise the difference between the curvature at a coordinate and the average curvature at that coordinate regularised by an elastic and a monotonic constraint. We use the same energy minimisation approach when imposing shape correspondence.

4.2. GPLVM Training

The type of mapping from the latent space to shape space is determined by the type of kernel used. In particular Lawrence [12] shows that with a linear kernel, the latent space of a GPLVM is the result of a PCA decomposition on the data. By ignoring the posterior variance in equation (19), we are essentially training a shape model that is similar to the one used by Cootes et al. [4].

Hence, to train a non-linear shape model, we use an Auto Relevance Determination (ARD) kernel

$$\kappa_{\text{ard}}(\boldsymbol{t}, \boldsymbol{t}') = \phi^2 \exp\left(-\frac{1}{2}\sum_{q=0}^{Q-1} \alpha_q (t_q - t'_q)^2\right) \tag{21}$$

in a Bayesian GPLVM, which, when trained with Q set to a high value, allows us to choose the latent space dimension of the GPLVM by looking at the decay of the length-scales α_q [26]. Figure 5 shows the latent space along with the corresponding generated shape.

The kernel we use for the linear shape model is given by

$$\kappa_{\text{lin}}(\boldsymbol{t}, \boldsymbol{t}') = \phi^2 (\boldsymbol{t})^t \boldsymbol{t}'. \tag{22}$$

which we use to train a GPLVM. When using the linear kernel, the training latent space parameters $\mathbf{T}$ are simply given by the PCA decomposition of the data matrix as shown by Lawrence [12].

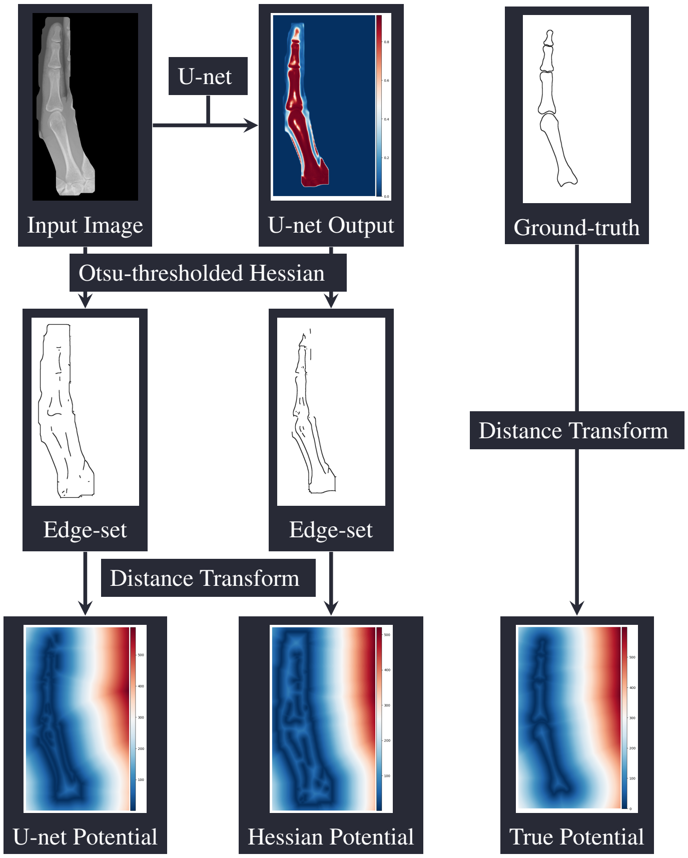

Figure 4. Creation of the three edge potential functions. In all cases, a distance transform is performed on the edge set. The U-net edge set is found by performing Otsu thresholding [14] on the hessian of the U-net bone predictions. The true edge potential is found by performing a distance transform on the true edge set.

4.3. Edge Discriminator

We train a convolution U-net [20] to act as a bone detector; details of the architecture and training for the U-net are provided in the supplemental material appendix C. Due to the incomplete annotation performed, segmentation masks could only be generated for the right index finger. To address this issue, images were cropped and masked. We modify the cross entropy loss so that only pixels around the masked area of interest contributed to the total loss. To turn the U-net output into an edge potential, we first use a Hessian Based edge detector [14] on the U-net output to find the edge set. We then find the distance transform of this edge set.

We also use the same Hessian based edge detector on the image directly to get an edge set which we distance transform. To assess the performance of these two edge potential functions, we create the true edge potential function by using the true outlines as the edge set in the distance transform. We show this process in Figure 4.

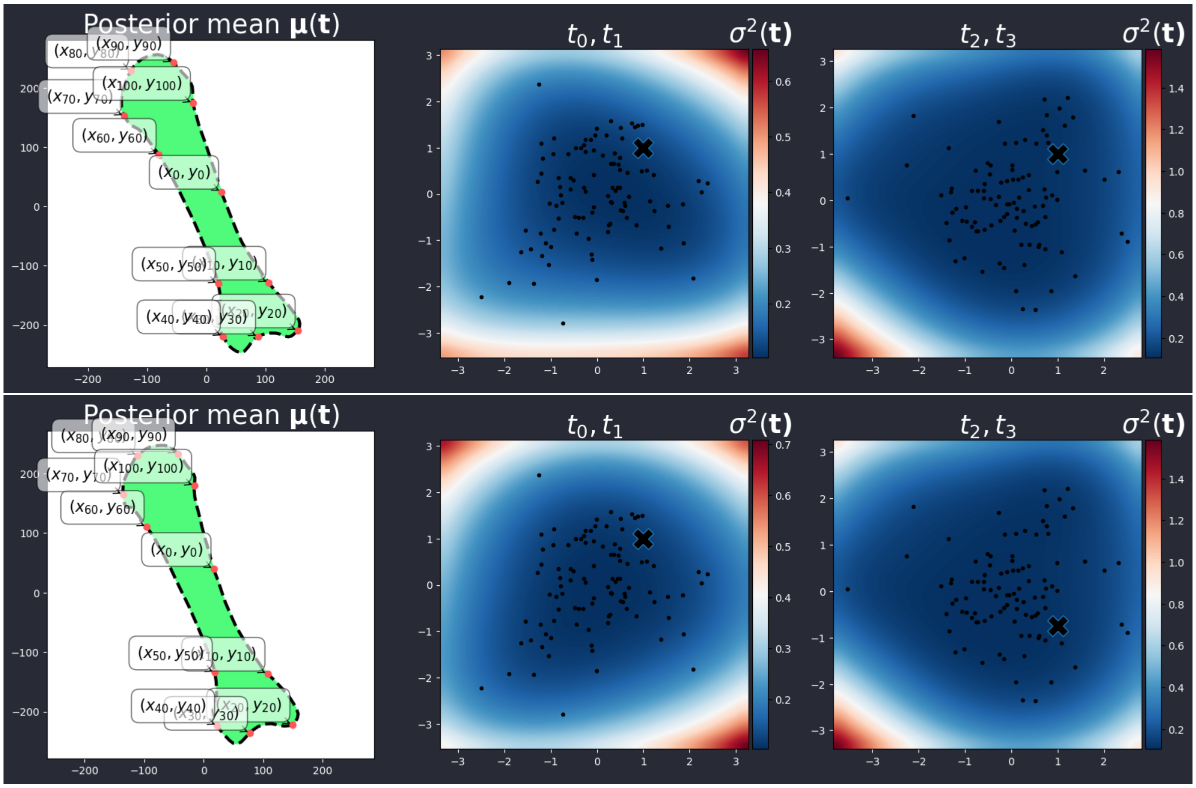

Figure 5. The plots shows a GPLVM generating two different shapes from two latent point positions $\boldsymbol{t}$. A latent space of dimension $Q = 4$ can be seen on the middle and right most plots on which the black cross shows the value $\boldsymbol{t}$ that generates a shape on the left-most plots. The latent plots show the posterior variance $\sigma^2(\boldsymbol{t})$ of the GPLVM.

4.4. Results and Discussion

Our results are shown in Figures 6 and 7 where we show the error for each bone when using different edge potential function and latent space dimensions. We show both the overall average error across the whole bone (Figure 6) as well as specifically the error in the estimation of the joint space width (Figure 7).

Latent Space Dimension: To understand the effect of the latent space dimension Q, we consider the accuracy when using the true bone outlines to build the edge potential function. We can see that the error decays as one increases the latent space dimension. This is because the higher dimensions are able to capture finer shape features; that is, higher Fourier modes in the expansion (20). We do not observe this when using the other discriminators as these fail to capture these high frequency features in the pixel domain.

Potential Function: The U-net potential does worse than the simple edge discriminator when detecting bones. However, it yields better results when measuring the joint width. As observed in Figure 4, U-net yields an edge potential function that is better at modelling the joint gap than the simple hessian edge discriminator. This is because the convolution kernels of U-net are extracting finer scale features that correspond to high dimensional derivative information.

The edge discriminators capture noise with short length scales which distorts the true shape of the outline as shown in Figure 4. The shape model overcomes this problem by instead capturing longer scale correlations in the pixel space through its prior.

Non-Linear Latent Space Mapping: The linear shape model shows worst accuracy when used alongside the Hessian and unet based potential functions. This difference is more pronounced when using the Hessian based potential function, which suggests that the non-linear kernel is more robust to noise present in the potential function.

5. Conclusion

We propose to use an ALSSM to segment bones from a hand in order to find the joint space. The model uses a non-linear GPLVM representation for shapes which are represented as discretised curves bounding an area of interest. We highlight the importance of the edge potential function

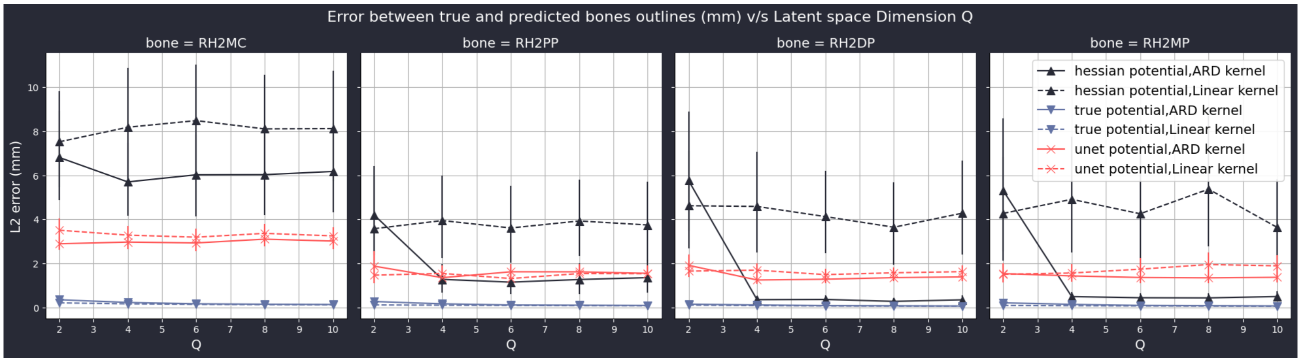

Figure 6. Results for the average L-2 error between the model generated bone outline and the true bone outline against the dimension Q of the GPLVM latent space in a 10-fold validation. We compare the performance of three edge potential functions; one built by a U-net discriminator (unet), one built using a hessian based edge finder (hessian) and finally by using the true bone outline to build the potential (true). We also investigate the effect of using a non linear kernel (ARD) and a linear kernel. The vertical lines represent error bars (standard deviation) of each mean.

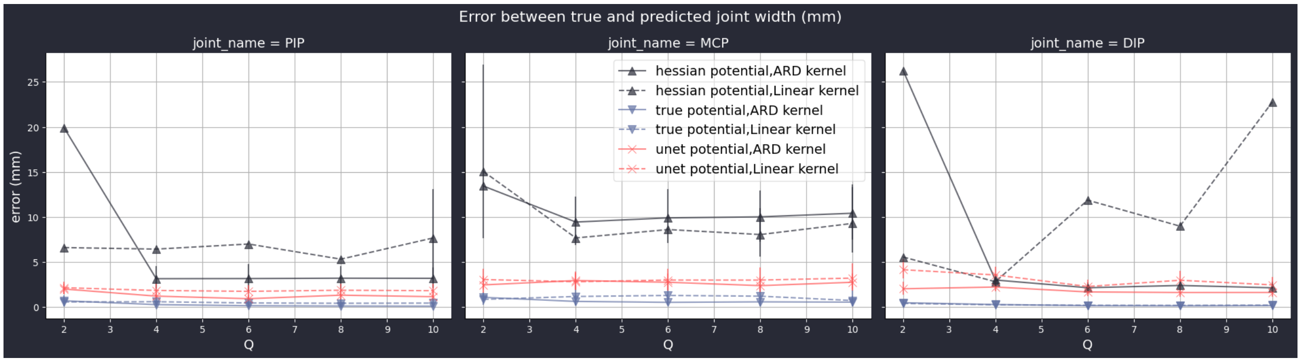

Figure 7. Results for the average error between the model generated joint space width and the true joint space width against the dimension Q of the GPLVM latent space in a 10-fold validation. We compare the performance of three edge potential functions; one built by a U-net discriminator (unet), one built using a hessian based edge finder (hessian) and finally by using the true bone outline to build the potential (true). We also investigate the effect of using a non linear kernel (ARD) and a linear kernel. The vertical lines represent error bars (standard deviation) of each mean.

by showing that it is what guides the performance of the model in a 10-fold validation result. Particular emphasis was laid on the full Bayesian Marginalisation of our model. The method we propose allows one to control the error obtained from the approximation in section 3.3 and captures the uncertainty in the predictions from the model which can be used to support down-stream clinical decisions.

Importantly, using the automatic alignment approach, we are able to make accurate estimates of the JSW without requiring detailed annotation from clinicians who would need to put all landmarks in correspondence manually, which would result in a large increase in the time and expense taken to label an x-ray.

Acknowledgements

This work has been supported by the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa, EP/L015684/1), the UKRI CAMERA Research Centre (EP/M023281/1 & EP/T014865/1), the UK National Health Service and the Royal Society.

References

[1] Anna S Antony, Andrew Allard, Adwaye Rambojun, Christopher R Lovell, Gavin Shaddick, Graham Robinson, Deepak R Jadon, Richard Holland, Charlotte Cavill, Eleanor Korendowych, et al. Psoriatic nail dystrophy is associated with erosive disease in the distal interphalangeal joints in psoriatic arthritis: a retrospective cohort study. *The Journal of rheumatology*, 46(9):1097–1102, 2019.

[2] Neill DF Campbell and Jan Kautz. Learning a manifold of fonts. *ACM Transactions on Graphics*, 33(4), 2014.

[3] Tim F Cootes, Mircea C Ionita, Claudia Lindner, and Patrick Sauer. Robust and accurate shape model fitting using random forest regression voting. In *European Conference on Computer Vision*, pages 278–291. Springer, 2012.

[4] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[5] Alessandro Di Martino, Erik Bodin, Carl Henrik Ek, and Neill DF Campbell. Gaussian process deep belief networks: A smooth generative model of shape with uncertainty propagation. In *Asian Conference on Computer Vision*, pages 3–20. Springer, 2018.

[6] Yangqin Feng, Daniel Lighter, Lei Zhang, Yan Wang, and Hamid Dehghani. Application of deep neural networks to improve diagnostic accuracy of rheumatoid arthritis using diffuse optical tomography. *Quantum Electronics*, 50(1):21, 2020.

[7] Yinghe Huo, Koen L Vincken, Max A Viergever, and Floris P Lafeber. Automatic joint detection in rheumatoid arthritis hand radiographs. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 125–128. IEEE, 2013.

[8] Matthias Kirschner, Meike Becker, and Stefan Wesarg. 3d active shape model segmentation with nonlinear shape priors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 492–499. Springer, 2011.

[9] Julia Kruger, Jan Ehrhardt, and Heinz Handels. Probabilistic appearance models for segmentation and classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1698–1706, 2015.

[10] Georg Langs, Philipp Peloschek, and Horst Bischof. Asm driven snakes in rheumatoid arthritis assessment. In Josef Bigun and Tomas Gustavsson, editors, *Image Analysis*, pages 454–461, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[11] Georg Langs, Philipp Peloschek, Horst Bischof, and Franz Kainberger. Model-based erosion spotting and visualization in rheumatoid arthritis. *Academic Radiology*, 14(10):1179–1188, 2007.

[12] Neil Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *The Journal of Machine Learning Research*, 6:1783–1816, 2005.

[13] Claudia Lindner, Shankhar Thiagarajah, J Mark Wilkinson, Gillian A Wallis, Timothy F Cootes, arcOGEN Consortium, et al. Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE transactions on medical imaging*, 32(8):1462–1472, 2013.

[14] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[15] Victor Adrian Prisacariu and Ian Reid. Nonlinear shape manifolds as shape priors in level set segmentation and tracking. In *CVPR 2011*, pages 2185–2192. IEEE, 2011.

[16] P Rahman, DD Gladman, RJ Cook, Y Zhou, G Young, and D Salonen. Radiological assessment in psoriatic arthritis. *Rheumatology*, 37(7):760–765, 1998.

[17] Adwaye Rambojun. *Automatic scoring of X-rays in Psoriatic Arthritis*. PhD thesis, University of Bath, 5 2020.

[18] Janick Rohrbach, Tobias Reinhard, Beate Sick, and Oliver Dürr. Bone erosion scoring for rheumatoid arthritis with deep convolutional neural networks. *Computers & Electrical Engineering*, 78:472–481, 2019.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[20] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

[21] Olga Schenk, Yinghe Huo, Koen L Vincken, Mart A van de Laar, Ina HH Kuper, Kees CH Slump, Floris PJG Lafeber, and Hein BJ Bernelot Moens. Validation of automatic joint space width measurements in hand radiographs in rheumatoid arthritis. *Journal of medical imaging*, 3(4):044502, 2016.

[22] John T Sharp, Gilbert B Bluhm, Andrew Brook, Anne C Brower, Mary Corbett, John L Decker, Harry K Genant, J Philip Gofton, Neal Goodman, Arvi Larsen, et al. Reproducibility of multiple-observer scoring of radiologic abnormalities in the hands and wrists of patients with rheumatoid arthritis. *Arthritis & Rheumatism*, 28(1):16–24, 1985.

[23] U. Snekhalatha and M. Anburajan. Dual tree wavelet transform based watershed algorithm for image segmentation in hand radiographs of arthritis patients and classification using bpn neural network. In *2012 World Congress on Information and Communication Technologies*, pages 448–452, 2012.

[24] William Taylor, Dafna Gladman, Philip Helliwell, Antonio Marchesoni, Philip Mease, and Herman Mielants. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 54(8):2665–2673, 2006.

[25] William Tillett. Oxford textbook of psoriatic arthritis: Plain radiography. In *Oxford Textbook of Psoriatic Arthritis*, pages 147–154. Oxford University Press, 2019.

[26] Michalis K Titsias and Neil D Lawrence. Bayesian Gaussian process latent variable model. In *International Conference on Artificial Intelligence and Statistics*, pages 844–851, 2010.

[27] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In *ICML*, volume 1, page 4, 2016.

[28] Martijn van de Giessen, Sepp de Raedt, Maiken Stilling, Torben B. Hansen, Mario Maas, Geert J. Streekstra, Lucas J. van Vliet, and Frans M. Vos. Localized component analysis for arthritis detection in the trapeziometacarpal joint. In Gabor Fichtinger, Anne Martel, and Terry Peters, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, pages 360–367, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[29] DMFM Van der Heijde, H PAULUS, and P SHEKELLE. How to read radiographs according to the Sharp/van der Heijde method. Discussion: Heterogeneity in rheumatoid arthri-

tis radiographic trials. Issues to consider in a metaanalysis. *Journal of rheumatology*, 27(1):261–263, 2000.

[30] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2097–2106, 2017.

[31] S Wassenberg, V Fischer-Kahle, G Herborn, and R Rau. A method to score radiographic change in psoriatic arthritis. *Zeitschrift für Rheumatologie*, 60(3):156–166, 2001.