Understanding Uncertainty Maps in Vision with Statistical Testing (Supplement)

1. Theoretical justifications

Lemma 1. *Let $Z(s)$ be a UGRF, and $\Lambda(s) = Var(\dot{Z}(s))$. The following properties are satisfied, [3]:*

1. *If Γ is spatially constant non-singular matrix, then $L_d(S, \Gamma'\Lambda\Gamma) = L_d(\Gamma^{-1}S)$.*
2. *If $\Gamma = \gamma I_{D\times D}$, then $L_d(\Gamma^{-1}S, \Lambda) = \frac{1}{\gamma^d} L_d(S, \Lambda)$.*
3. *If $S' = \Gamma S$, then $\Lambda(S') = \Gamma' Var(\dot{Z}(S))\Gamma$.*

Theorem 3.2. (In the main paper). *The domain S of the GRRF F can be warped via a one-to-one smooth transformation Γ to a domain S' without fundamentally changing the problem, namely: $\mathbb{P}\left(\max_{s'\in S'} F(s') \geq t\right) = \mathbb{P}\left(\max_{s\in S} F(s) \geq t\right)$.*

Proof. Since Γ is a one-to-one mapping, let $S' = \Gamma^{-1}S$ and $S = \Gamma S'$. Since F is a Gaussian Related RF, we consider Z as an underlying UGRF, and denote Λ' and Λ as variance of spatial derivatives of Z on S' and S, respectively. We notate Euler density of F as ρ. To show that with warping we do not change the problem, we need to show the following:

$$\mathbb{P}\left(\max_{s'\in S'} F(s') \geq u\right) = \mathbb{P}\left(\max_{s\in S} F(s) \geq u\right).$$

We start with Application of GKF Theorem (from the main paper):

$$\begin{aligned}
&\mathbb{P}\left(\max_{s'\in S'} F(s') \geq u\right)\\
\approx &\sum_d L_d\left(S', \Lambda'\right)\rho_d(u)\\
= &\sum_d L_d\left(\Gamma^{-1}S, \Lambda'\right)\rho_d(u)\\
= &\sum_d L_d\left(S, \Gamma'\Lambda'\Gamma\right)\rho_d(u)\\
= &\sum_d L_d\left(S, \Gamma'\text{Var}\left(\dot{Z}(S')\right)'\Gamma\right)\rho_d(u)\\
= &\sum_d L_d\left(S, \text{Var}\left(\dot{Z}(\Gamma S')\right)\right)\rho_d(u)\\
= &\sum_d L_d\left(S, \text{Var}\left(\dot{Z}(S)\right)\right)\rho_d(u)\\
= &\sum_d L_d\left(S, \Lambda\right)\rho_d(u)\\
= &\sum_d \mathbb{P}\left(\max_{s\in S} F(s) \geq u\right)
\end{aligned}$$

□

Remark 1. *For u_Z, s.t. $\mathbb{P}\left(Z_{\max} \geq u_Z | H_0\right) = 0.05$, $\rho_d^Z(u_Z) > 0$ for $d = 0, \ldots, 3$. Note, it is actually true for $u << u_Z$, but we only need u_Z case. We focus on $d = 0, \ldots, 3$, because it is the most common case in practical application, and corresponds to the dimension of the Search region/Domain S. Usually in computer vision we only focus on 2d and 3d images and do not consider domains with dimension > 3.*

Proof. It is easy to see that from the form of $\rho_d^Z(u)$.

$$\rho_0^Z(u) = \int_u^\infty \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-u^2/2} du$$

$$\rho_1^Z(u) = \frac{\lambda^{\frac{1}{2}}}{2\pi} e^{-u^2/2}$$

$$\rho_2^Z(u) = \frac{\lambda}{(2\pi)^{\frac{3}{2}}} e^{-u^2/2} u$$

$$\rho_3^Z(u) = \frac{\lambda^{\frac{3}{2}}}{(2\pi)^2} e^{-u^2/2}\left(u^2 - 1\right)$$

The EC densities $\rho_d^Z(u)$ depend on a single parameter λ, the *roughness* of the random field. We can see that for $u > 1$ and $d = 0, \ldots, 3$, $\rho_d^Z(u) > 0$. And in a case of 2d GRF, common in vision, $u > 0$ is enough to get $\rho_d^Z(u) > 0$. □

Theorem 3.3. (In the main paper). *Consider Gaussian Related RF $F(S)$ on domain S with Euler densities $\{\rho_d^F(u)\}$, and underlying UGRF $Z(S^Z)$ on domain S^Z with Euler densities $\{\rho_d^Z(u)\}$. Assume, that both Euler densities $\{\rho_d^F(u)\}$ and $\{\rho_d^Z(u)\}$ are defined on the same domain $u \in U$ and $\max\{\rho_d^F(u)/\rho_d^Z(u)\} \leq 1$. Then, by finding a one-to-one transformation Γ, such that $S = \Gamma S^Z$ and $S^Z = \Gamma^{-1}S$, and selecting threshold u^*, such that $\mathbb{P}\left(\max_{s\in S^Z} Z(s) \geq u^*\right) = 0.05$, guarantees that $\mathbb{P}\left(\max_{s\in S} F(s) \geq u^*\right) \leq 0.05$.*

Proof. Let $\rho_d^F(u) = c_d(u)\rho_d^Z(u)$. Given the Remark 1, consider two cases with respect to sign of $\rho_d^F(u)$:

1. If $\rho_d^F(u) < 0$, then $c_d(u) < 0$, then $\rho_d^F(u) < \max\{c_d(u)\}\rho_d^Z(u)$.
2. If $\rho_d^F(u) \geq 0$, then $c_d(u) \geq 0$, then $\rho_d^F(u) \leq \max\{c_d(u)\}\rho_d^Z(u)$.

That is, for any sign of $\rho_d^F(u)$, we have $\forall d$, $\rho_d^F(u) \leq c\rho_d^Z(u)$, where $c = \max\{c_d(u)\}$. And there $\exists j$, such that $\rho_j^F(u) = c\rho_j^Z(u)$. Assuming that $\max\{\rho_d^F(u)/\rho_d^Z(u)\} \leq 1$, we get $c \leq 1$.

We start with an application of GKF Theorem (from the main paper):

$$\begin{aligned}\mathbb{P}\left(\max_{s\in S} F(s) \geq u\right) &\approx \sum_d L_d(S, \Lambda^F)\rho_d^F(u)\\ &\leq c\sum_d L_d(S, \Lambda^F)\rho_d^Z(u)\\ &\leq L_d(S, \Lambda^F)\rho_d^Z(u)\end{aligned}$$

Applying Thm. 3.2, if we find a warping Γ, s.t. $S = \Gamma^{-1}S^Z$ and $S^Z = \Gamma S$, then

$$\begin{aligned}&= L_d(S^Z, \Lambda^Z)\rho_d^Z(u)\\ &= \mathbb{P}\left(\max_{s\in S^Z} F(s) \geq u\right) \qquad \square\end{aligned}$$

2. Bootstrap

In the method section of the main paper ('Uncertainty between the models'), we briefly attributed to an application of bootstrap technique to approximate the distribution of test statistics $\{F_i\}_{i=1}^N$ under the null H_0 and conduct a hypothesis test, by checking if observed statistics falls into the rejection region, defined by corresponding u_F.

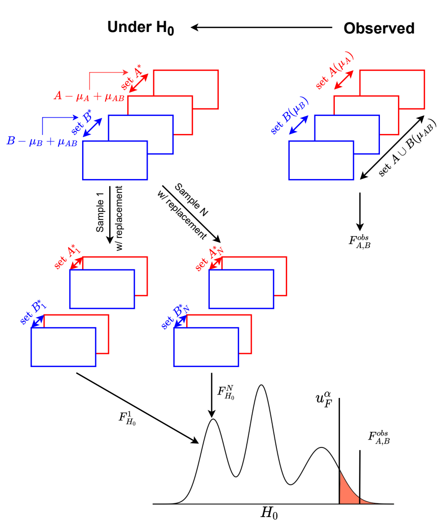

Figure 1. Illustration of application bootstrap technique to generate distribution of statistics F under H_0.

The bootstrap procedure for approximation of distribution of test statistics under H_0 is presented in Algorithm 1 and illustrated in Fig. 1.

Algorithm 1 Bootstrap algorithm for approximation of test statistics distribution under H_0

Input: Uncertainty masks for models A and B, $A_x(s)$ and $B_x(s)$ correspondingly, where s represents pixel
Output: Distribution of test statistics F under H_0

Require: Test statistics $F(x; y)$ between masks x and y

1: Compute observed statistics $F^{obs}_{A;B} := F(A_x; B_x)$
2: Generate 2 data sets of uncertainty masks under H_0:
$A^* = A$ group + overall and
$B^* = A$ group + overall
3: **for** i 1 to N **do**
4: Sample with replacement sets A_i^* and B_i^* from the combined set A^* [B^*
5: Compute corresponding $F^i_{H_0} := F_{A_i^*;B_i^*}$
6: **end for**
. Generated $fF^i_{H_0}g^N_{i=1}$ approximates the distribution of test statistics $F(x; y)$ under H_0

Note: Algorithm 1 requires definition of test statistics $F(x, y)$. For example, it can be a common student statistics $F_{A,B}(s) = \overline{(A_x(s) - B_x(s))} / \sigma(\overline{A_x(s) - B_x(s)})$, where σ is an estimate of standard deviation and $\bar{x}$ represents the sample mean.

The approximation of the distribution of test statistics $\{F_i\}_{i=1}^N$ under the null hypothesis H_0 allows us to find a threshold u_F, such that $\mathbb{P}(F_{\max} \geq u_F | H_0) = \alpha$. Thus, if observed test statistics $F^{obs}_{A,B}$ is higher than this threshold u_F (or in our case significant mask $\widehat{\mathcal{M}}_F$ contains significant pixels), then we reject H_0 in favour of H_A.

3. Description of the Warping NODE Networks

3.1. Warping NODE

Learning the warping $\Phi(S)$ defined in equation 1.

$$\frac{d\Phi_t}{dt} = V(\Phi_t), \tag{1}$$

As we mentioned, to model warping Φ_t we follow a common in vision technique [1, 2]. The main idea is to model warping as target coordinates, from which we sample (or move to). That is, for input image with dimension $h \times w$, Φ_t is of dimension $h \times w \times 2$, where 2 appears because the dimension of image is 2. Since we model warping Φ_t as NODE, the modelled derivative $V(\Phi_t)$ have to be of the same size $h \times w \times 2$. Since it is important to generate different warping according to the input x, we model $V(\Phi_t)$ conditioning on image x, as $V(\Phi_T, x)$. We do this using U-Net, demonstrated in Fig. 2 (Top) with number of layers/channels depending on a problem.

3.2. Discriminator

Discriminator is used to minimize Wasserstrein distance, and presented in Fig. 2 (Bottom).

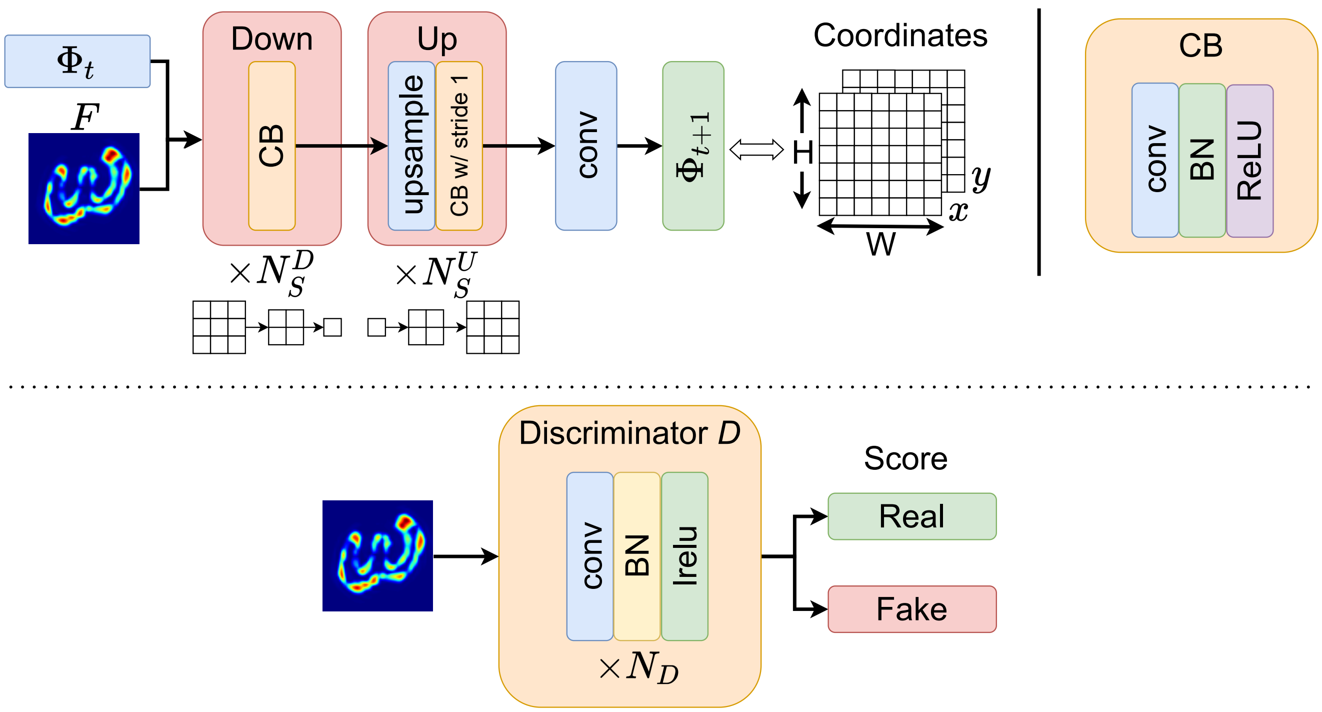

Figure 2. *Top:* Warping Neural ODE part, which is used to generate the warping $_{t+1}$ at time point $t + 1$. *Bottom:* Discriminator D, main goal of which is to differentiate between warped RF $\hat{F}$ and observed GRF Z. N stands for number of convolution blocks.

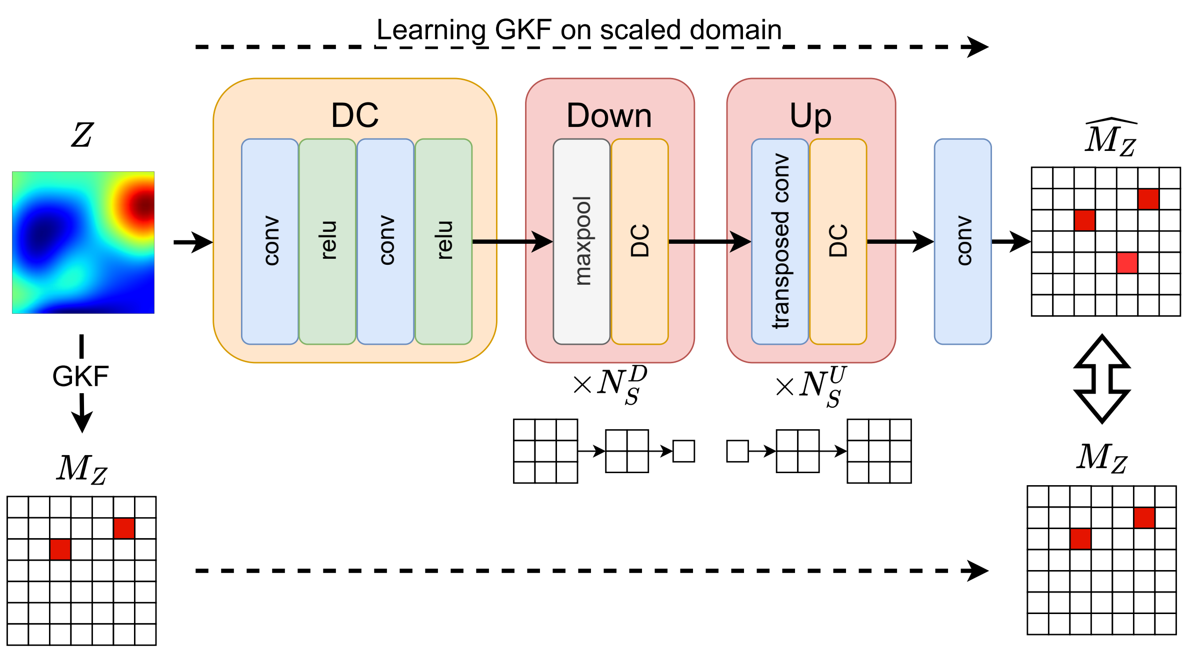

Figure 3. Segmentation network, the main goal of which is to learn the GKF theorem on a scaled domain and derive the significance mask $\hat{M}_Z$ and later $\hat{M}_F$. The implementation is based on U-net structure. DC block corresponds to double convolution block, where first convolution changes number of channels and second preserves it.

3.3. Notes on assumptions of the Theorem 3.3 and GKF

One of the assumption of the Theorem 3.3 is that $\rho^F(s)$ and $\rho^Z(s)$ are defined for the same domain $s \in S$. In other words, since we are focusing on the warping, rather than Generative models, we are not allowed to generate pixels with completely new color values, but only warp (make wider, change order) observed pixels. For this reason, we scale both observed RF F and GRF Z to the same scale, e.g. [0, 1]. Since the scale is changed, we cannot directly use significant thresholds u_Z, obtained from GKF. We have to properly scale it, corresponding to the scaling used for F and Z. Instead, we use a "smart" scaling. To derive the threshold, corresponding to the scaled u_Z, we train (in a supervised manner) a segmentation network on (scaled GRF Z; significant mask M_Z, derived before scaling based on u_Z), represented on Fig. 3. Then resulted network is applied to the warped image $\widehat{F}$ to derive $\mathcal{M}_F$.

3.4. Notes on generating GRF

The main idea of our method is based on mapping an uncertainty map to the isotropic GRF, for which we can derive significant region based on GKF theorem. To generate an isotropic GRF, it is necessary to provide 1) dimensions of a RF, which corresponds to the dimension of the input uncertainty map, 2) covariance function: $C(t) = \exp\left(-\beta||t||^2\right)$ with the proper scale β. Note that scale β directly affects the threshold u_Z for which elements (in our case pixels) of GRF is considered to be significant, through the computation of LKC (or EEC). We consider scale β as a hyper parameter, which impacts how easy it is to find a warping from uncertainty map to GRF, and mainly depends on an uncertainty map size. However, given the scale it is important to recompute a threshold u_Z to generate a significant mask $\mathcal{M}_Z$ for GRF.

4. Experiments

4.1. Synthetic

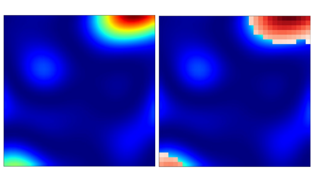

Figure 4. (Left) GRRF, ^{2} with 5 degrees of freedom and (Right) overlapped significance mask $\mathcal{M}_F$ derived by our model.

We show our model learns correct significant region, preserving power and limiting type-1 error rate by a significance level. We generate 1000 samples of Gaussian RF Z and 3 sets of the 1000 samples of Chi-squared RF F, with the following degrees of freedom: 1, 3, and 5. All sets were generated to satisfy the conditional that $\mathbb{P}(Z_{\max} \geq u_Z | H_0) = 0.05$ and $\mathbb{P}(F_{\max} \geq u_F | H_0) = 0.05$. Since there is a closed form solution to compute $\mathbb{E}\{\phi(A_u)\}$ for F, we can obtain a ground-truth significance mask $\mathcal{M}_F$. Therefore, we can evaluate the performance of our model by comparing generated $\mathcal{M}_F$ with a real (ground truth) $\mathcal{M}_F$. Tab. 1 provides achieved power and type-1 error on the learned test. We see that for all cases power of test is higher than 80%, while preserving type-1 error less than 0.05.

	χ^2_1	χ^2_3	χ^2_5		χ^2_1	χ^2_3	χ^2_5
Power	0.83	0.83	0.87	Type-1	0.02	0.02	0.02

Table 1. The achieved power and type-1 error of our generalized test, performed on three different GRRF, i.e. ^{2} with varying degrees of freedom.

4.2. Longitudinal Brain Image data, ADNI.

Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu). It contains gray matter probability masks from the T-1 weighted MR images of size $105 \times 127 \times 105$ per subject at 3 time steps. The subjects are divided into two groups: diagnosed with Alzheimer's disease (AD: 377 subjects) and healthy controls (CON: 152 subjects). Results mentioned in the main paper are displayed in Fig. 5.

References

[1] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007. 2

[2] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28:2017–2025, 2015. 2

[3] Jonathan E Taylor et al. A gaussian kinematic formula. *Annals of probability*, 34(1):122–158, 2006. 1