

1 We thank the reviewers for their valuable feedback, and we address raised questions and comments below.

2 **R1: Discussion of limitations, repetitive patterns.** Thank you for pointing to the interesting question regarding
3 regular textures. Indeed Perlin noise based techniques by design cannot synthesize repeating textures, since the texture
4 is the result of sampling an underlying infinite noise field, designed to capture randomness in natural objects. It is
5 therefore well suited for natural 3D textures. We will discuss these limitations explicitly in our revised text.

6 **R2: Machine learning contributions.** A main novelty is our proposed loss function, with *discriminator features for*
7 *style loss* leading to better results than existing approaches (GAN loss, VGG style loss) for texture synthesis. Our
8 loss enables training *a conditional GAN using an unconditional discriminator*. This is crucial in texture synthesis
9 and related fields: This not only relaxes constraints on the discriminator (therefore providing better gradients to the
10 generator), but it also effectively avoids practical difficulties with a conditional discriminator. This can be imagined,
11 e.g., considering an exemplar image where the texture in the top left differs from the texture at the bottom right. We
12 evaluated the conditional alternative extensively and we will emphasize this aspect in more detail in the text. We expect
13 our novel loss to be applicable beyond texture synthesis; potentially in related fields such as style transfer.

14 **R2: Rendering during training.** In this work we learn only diffuse textures. An explicit rendering step is thus omitted
15 and the loss is computed directly on the synthesized slice. After training, as an application demo we use the learned
16 textures in a rendering framework (lines 265-267) under various lighting conditions. We will clarify this in the text.

17 **R2: Use of word "frequency".** By frequency we refer to how we sample the noise fields (which we do in a periodic
18 manner), not the noise signal itself. We will better define this in the text.

19 **R2: Discussion of references [17,18].** We mentioned these references in Section 2 (line 100-101). Similarity of these
20 approaches to ours is a noise field being transformed to a texture, enabling infinite and seamless synthesis. However, in
21 [17,18] the transformation is parametrized with CNNs unlike ours. While this is efficient in 2D, it is not feasible in
22 3D. The most important difference is their restriction to 2D as well as our novel loss function. We will discuss these
23 references comparatively and add a comparison to the supplemental material.

24 **R2: Relevance of Section 3.3 on texture extrapolation.** We believe that our extrapolation strategy is important in
25 practice since it reduces training time substantially for adding any new samples to a dataset, which is common in
26 production. Although we would prefer to keep this section in the main text, if there is consensus that it does not
27 strengthen the submission, we will gladly move it to supplementary material.

28 **R2: Complexity of shown examples.** Many common texture synthesis examples are not 3D volumetric materials. Our
29 application domain of 3D textures somewhat limits our choice of examples, e.g., to volumetric materials like stone and
30 wood. We will include additional and more diverse examples in the supplemental material.

31 **R2: Training patch size.** The training patch size (which defines the receptive field) can be chosen arbitrarily, i.e.,
32 our system can be trained on larger patches (256^2 or higher). Such choice determines the range of frequencies in the
33 training data captured by our network. We will mention this in the text.

34 **R3: Not self-contained.** Thank you for pointing this out. We will gladly include the missing details in the text.

35 **R3: Novelty in replacing VGG with discriminator.** It is true that perceptual loss on discriminator features has been
36 proposed earlier, we will discuss this explicitly in the revised text. This concept has been applied in image reconstruction
37 and image translation tasks, where an input is mapped to its ground truth. However, we do not aim to reconstruct the
38 content of our input, but to reproduce its style. Thus we propose a novel style loss based on discriminator features.

39 **R4: Design choice to split frequencies.** With this we are able to reduce the training time by roughly 20% without
40 noticeable degradation of output image quality. We will gladly include an ablation experiment.

41 **R4: Evaluation of diversity.** We would like to emphasize that our model provides diverse outputs, by construction,
42 similarly to [5] (due to sampling and combination of infinite noise fields). It is true that we do not quantitatively evaluate
43 diversity in an isolated manner; however, we demonstrate diversity qualitatively by synthesizing large stripes, several
44 times larger than the training exemplars. In addition, we report FID, which was shown in the literature to correlate
45 well with both quality and diversity (although not a perfect metric). We believe that our quantitative evaluation with
46 synthesis of 50 slices that are significantly larger than the receptive field/training patch size demonstrates diversity.

47 **R4: Comparisons to 2D texture CNNs ([a,b,c]).** We already discussed [a,b] in the original submission. Thank you
48 for pointing out reference [c], which we will gladly include. A visual comparison to demonstrate the limitations of 2D
49 approaches in comparison to ours is an excellent idea, thank you for suggesting this. In the supplementary material, we
50 will gladly provide such a comparison, which will strengthen our manuscript.

51 **R4: Generalization w.r.t. Gram matrix distance.** Thank you for your suggestion. This is indeed an interesting
52 experiment. We will investigate such relationship and report our findings in the revised version.