We thank all reviewers for careful reading of our paper and their positive, encouraging, constructive, and in-depth feedback. In the following, we will address the concerns raised in the reviews. Any clarifications given in this author response will be included in the revised version.

Reviewer #1. ***Contribution and impact.*** Our main contribution is the approximation of a highly non-trivial function by means of deep learning where model-driven approaches have been cumbersome and computationally complex. In that, we decrease the computational requirements substantially. While we do use a well known architecture as the backbone to our approach, we take a step forward in demonstrating intrinsic properties of the non-linear decomposition (e.g. one-homogeneity) are retained and our approach generalises very well to eigenfunctions of the TV subdifferential (very different from the training set). The spectral TV decomposition can be seen as a non-linear analogue to the Fourier transform, and we aim to in some sense obtain an analogue to Fast Fourier Transform (FFT). The availability of a fast computational method is one of the reasons for the success of the Fourier transform in signal and image processing. Fast methods for non-linear spectral decomposition have the potential to become an equally important tool for image and data analysis. ***Applications for spectral TV decomposition.*** Non-linear spectral decomposition has many successful applications in imaging including denoising [35], texture extraction and separation [6, 23], fusion [4, 46, 20] or segmentation [43] to mention a few. Especially in image fusion applications, spectral TV decomposition is able to overcome challenges in relation to edge, and detail preservation where other methods fail [46]. The use of spectral TV in large-scale applications is hampered by its computational complexity, which is the issue we address in this paper. The advantage and strength of using spectral TV in applications specifically in comparison to, e.g., fully learned approaches is the solid theoretical background. The spectral TV transform is based on the L1 norm and hence is robust, it is convex, edge preserving and does not require parameter tuning. It is mathematically very well understood giving a fundamental non-linear representation, and therefore its use leads to more sound and explainable methods. ***Introduction and background.*** The reviewer is concerned that the introduction and background are not accessible enough in light of spectral TV decomposition being relatively unknown to the larger community. We will address this issue by adding more high level explanations along the lines of the above. However, we believe that the mathematical details given are necessary to not only understand the complexity of the original problem, but also to introduce the properties known to hold true for spectral TV (such as one-homogeneity) that are later shown to be retained by TVspecNET.

Reviewer #3. ***Approximating function with deep neural network.*** While we do approximate a function by a neural network, this function is very non-trivial: it is the second derivative of a solution to a highly non-linear and non-smooth PDE, producing a non-linear decompositon of an image. Our main contribution is to use a neural network to reproduce this non-trivial decomposition while retaining intrinsic properties and as a result decrease the computation time by several orders of magnitute. ***Fixed number of bands.*** As suggested by the reviewer, we trained the same network for a finer graded decomposition with more bands, namely 25 bands, and the results are similarly convincing and faithful (SSIM: 0.958, PSNR: 30.228). However, it is not necessary to retrain the network for a larger number of bands to obtain a finer graded decomposition of certain scales: due to the one-homogeneity one can shift bands in an image by multiplying it by a constant factor (factor <1 will shift larger structures to smaller bands, a factor >1 will do the reverse). Hence, we can represent the structures of interest at a finer scale through shifting without having to retrain the network. ***Additional information on DnCNN.*** To ensure clarity and completeness, we will elaborate on the network architecture DnCNN that was used as the basis for TVspecNET. ***Other issues.*** We thank the reviewer for pointing out the missing reference to Table 3 in Section 4.3 and the typo in line 215. We will rectify both issues for the revised version.

Reviewer #4. ***Clarification of our contribution.*** We thank the reviewer for the remarks on the relevance of our submission. We would like to clarify that we are solving a challenging and non-trivial problem. That is, we approximate the otherwise computationally expensive spectral TV decomposition as it involves solving a highly non-linear and non-smooth PDE, and achieve a much lower computational cost. ***Main relevance.*** The reviewer pointed out a statement in the broader impact section that upon reflection we see can be easily misunderstood. Instead of claiming the main relevance to be the approximation of the PDE, we intended to highlight that we compute a decomposition that can already be computed with classical methods (and hence the societal impact should be of no concern), but much faster. We will clarify this in the revised version.

Reviewer #5. ***Proposed method.*** We will take a number of steps to improve the description of our proposed methods to ensure a clear understanding of our central idea. The derivation of the spectral TV decomposition of an image is challenging and computationally expensive as it amounts to solving a highly non-linear and non-smooth PDE and taking the second temporal derivative. We use a neural network to approximate the decomposition at a considerably reduced cost and further show the properties of the analytic TV transform also hold true for TVspecNET. We will add more details about our basis architecture (DnCNN) and extend the high level description by an analogy between the Fourier Transform and spectral TV, and FFT and TVspecNET. ***Other issues.*** We thank the reviewer for pointing out the typo, we will do a full proof read of the revised version.