We thank the reviewers for their valuable and constructive feedback. We will incorporate the suggestions in the next revision. Our responses to some specific comments are below.

Common Comments:

- **Limitation of Lower Bounds**: We argue that our lower bounds are strong since they show that even in the case of a 1-D function (which is the simplest case for representation), the lower bounds match the upper bounds. A one dimensional function can be embedded in $\mathbb{R}^d$ by considering $g(x) := f(\langle e_1, x\rangle)$ and the lower bounds follow. A similar trick is used in Telgarsky's work which shows depth separation.

  This work focuses on dimension-free bounds and therefore the lower bounds are also given in terms of the 'Fourier norms' which are dimension-free. The Fourier norms may themselves depend on the dimension for specific function classes like in Example 1 of Section 2, but the dimension dependence (if at all) is through the Fourier norms.

- **Regarding the Function Spaces**: It is indeed the case that the function spaces $\mathcal{G}_K$ are not easily characterized in terms of standard function spaces. Some intuition has been given in Examples 1 and 2 of Section 2. Similar spaces are used in classic works of Barron et. al to show the first quantitative representation theorems for neural networks and such function classes are well investigated in the approximation theory literature. We will add more references to these works and further explain features of $\mathcal{G}_K$ like 1) uniform continuity and 2) density of a subspace of $\mathcal{G}_K$ in $L^p(\mathbb{R}^d)$. The optimality of our results suggests that the spaces $\mathcal{G}_K$ are natural in the study of neural networks.

Reviewer Specific Comments:

Reviewer 1: See items 1 and 2 in common comments. The reviewer claimed that the lower bounds are already known in the literature. We could not find the references to such results. We request the reviewer to share any relevant references regarding this.

- **Regarding Literature Review** [23] considers depth separation - i.e, they construct functions which are outputs of a D layer network with a small number of neurons but need a huge number of neurons to represent with fewer layers ($D' \ll D$). We use the basic idea of bounding oscillations to show a quantitative lower bound on accuracy when representing certain functions in the class $\mathcal{G}_K$.

  [22] uses a careful manipulation of the Parseval-Plancherel formula, radial symmetry and the (lack of) support of Fourier transforms of neural network outputs to obtain a specific distribution and a specific function under which 2 layer networks cannot approximate radial functions easily. This is different from the present work which obtains a sampling procedure via the Fourier distribution in order to show a strong approximation result.

  We will make these distinctions clear in the next version of our manuscript.

Reviewer 2:See item 1 in Common Comments.

Reviewer 3: See item 2 in Common Comments.

Reviewer 4:

- **Regarding Relevance of the Work**:

  We did not discuss the training aspects of neural networks because we only concentrated on the representation power of deep networks. Based on current literature, theoretically understanding the SGD based training of deep neural networks appears to be a very hard problem. Nevertheless, representation results are important and interesting, and the long line of work on representation power of neural networks shows continued interest in the topic. Precise and optimal bounds on the expressive power of neural networks are algorithm independent and fundamental properties of neural networks. We believe it is essential to understand these aspects in order to understand specific training algorithms like SGD. For instance, in the paper https://arxiv.org/pdf/2001.04413.pdf, a precise upper bound on representation power of kernels is used to show that 3 layer networks outperform 2 layer networks in certain learning tasks via SGD type algorithms.

- **Regarding computational experiments**: Computational experiments of this nature would unfortunately be misleading because they would compare training via SGD to a purely representation result. This, in our opinion, cannot be justified easily and does not contribute to the results established in the paper.