We thank the reviewers for their detailed feedback and suggestions. We are happy to see that R2/R3 appreciate our "unified" description and implementation of E(2)-equivariant CNNs in an "umbrella framework" and their "extremely thorough" evaluation. As R1 is concerned about the significance and novelty of our contributions in general, we will clarify and accentuate our contributions as summarized below.

R1 - Incremental work: Indeed the idea of E(2)-equivariant CNNs is not new and we build on the well known framework of steerable CNNs [1,2,3,4,5]. However, the *theoretical* framework (Sec. 2.1-2.3) was so far instantiated only for a few choices of groups and representations [1,2,5] while a fully general solution of this important model class was still missing. The group restriction operation (exploiting local symmetries), the possibility to build hybrid models (operating on different field types simultaneously) and a large-scale benchmark study are also novel to the literature.

R1 - Significance of the general solution of the kernel constraint:

- Our general solution yields a unified description and implementation of a large body of E(2)-equivariant models [1,6,7,8,9,10,12,13,22,23,24] whose specific interrelations were so far left unclear.
- While these models were proven to be equivariant, most of them lacked a proof of implementing the most general equivariant mapping between the corresponding field types. By proving the completeness of our general solution we add these missing pieces. In the case of Harmonic Networks [12] we found that the solution space was incomplete.
- Our solution yields an off-the-shelf method for convolutions between arbitrary field types which were not being considered before. It describes e.g. convolutions from regular fields to vector fields which might become important e.g. for the processing of optical flow or wind fields. A unified implementation enables application focused researchers to start using equivariant models without needing to take the hurdle of deriving the appropriate mappings.
- Beyond planar CNNs, the solved kernel space constraint applies exactly to Gauge Equivariant CNNs [5] on 2-dim manifolds. In contrast to the plane $\mathbb{R}^2$, the structure group of Riemannian manifolds can in general *not* be reduced further than O(2). A general solution of the kernel space constraint is therefore strictly necessary for such models.

We agree with R1 that these points should be presented more clearly. We updated the paper accordingly and further added two sections discussing the incompleteness of Harmonic Networks and the relations to Gauge CNNs.

R1 - Significance of empirical evaluation: We want to point out that the significance of the conducted experiments goes beyond "showing that symmetries in the dataset are important". In particular:

- We benchmark 44 models (30 new ones) with different levels of equivariance, group representations and nonlinearities against each other (Tab. 9, supplementary). This experiment is of great importance for the field of equivariant deep learning since a direct comparison of the E(2)-equivariant model zoo was missing before our submission.
- The benchmark study was repeated on three transformed MNIST versions (standard, SO(2) and O(2)) which was not being done before. This allows to separate the contributions of global and local symmetries to the final performances.
- We evaluate the new group restriction at different depths on MNIST-rot (Tab. 10, suppl.) and on CIFAR-10/100 (Tab. 3) to support our hypothesis that exploiting local symmetries is beneficial even if not being present globally.
- Two (now three) models on MNIST-rot are presented which beat the previous SOTA (Tab. 2).
- Our CIFAR-10/100 models, trained with the auto-augment (AA) policy of [21], significantly outperform the AA baseline (Tab. 3). This shows the benefit of equivariance even when powerful, task-adapted augmentation is used.

R1 - paper organization: We presented the general theory of steerable CNNs (Subsec. 2.2/2.3) and our new contributions (Subsec. 2.4-2.8) together in our theory section 2 to give a comprehensive and self-contained exposition. In the final version we will clearly separate the new contributions from previous work.

R2/R3 - background required: While we tried to keep the model definitions self-contained, some knowledge in group representation theory is required. Background material, including definitions of induced and restricted representations as requested by R3, are added to the supplementary. We will further try to simplify the paper in general.

R3 - details for only one particular group: Our results in Subsec. 2.5-2.8 apply to a whole family of (orthogonal) groups O(2), SO(2), D_N and C_N. The general solution strategy in Subsec. 2.4 applies generally to any finite dimensional unitary representation and thus gives a clear roadmap for deriving general solutions for other groups.

R3 - improvements despite auto-augment (AA): Augmentation typically applies geometric transformations globally while GCNNs exploit local symmetries [5]. While GCNNs exploit symmetries by design, standard CNNs need to *learn* augmented samples explicitly. The hypothesis space and parameter cost of GCNNs are thus greatly reduced which leads to superior performances [1,2,5-10,12-14,17,18,22]. See Sec.2 and Sec.6, Par.1 in [7] for a more technical explanation.

R3 - overfitting: To prevent from overfitting we did not tune the hyperparameters in *any* of our experiments but simply replaced the non-equivariant convolutions of the baseline models with G-steerable convolutions.

R3 - new experiments: We are running new experiments on STL-10 to 1) confirm that our previous results generalize to higher resolutions and 2) investigate the improved sample complexity of GCNNs in a data ablation study. By replacing conventional with steerable convolutions we 1) significantly improved upon the previous (supervised) SOTA and 2) find our models to consistently outperform baselines on all dataset sizes tested so far (still gathering more samples).

An implementation on manifolds would indeed be interesting but would raise additional engineering questions which we believe to deserve their own publication. Our implementation on $\mathbb{R}^2$ allows for using a regular grid and therefore an investigation of steerable convolutions without the additional complications arising from discretizations of 2-manifolds.

We further found a bug in our CIFAR-10/100 models which now perform even (significantly) better.