UNIVERSAL APPROXIMATION THEOREM FOR EQUIVARIANT MAPS BY GROUP CNNS

Anonymous authors
Paper under double-blind review

ABSTRACT

Group symmetry is inherent in a wide variety of data distributions. Data processing that preserves symmetry is described as an equivariant map and often effective in achieving high performance. Convolutional neural networks (CNNs) have been known as models with equivariance and shown to approximate equivariant maps for some specific groups. However, universal approximation theorems for CNNs have been separately derived with individual techniques according to each group and setting. This paper provides a unified method to obtain universal approximation theorems for equivariant maps by CNNs in various settings. As its significant advantage, we can handle non-linear equivariant maps between infinite-dimensional spaces for non-compact groups.

1 INTRODUCTION

Deep neural networks have been widely used as models to approximate underlying functions in various machine learning tasks. The expressive power of fully-connected deep neural networks was first mathematically guaranteed by the universal approximation theorem in Cybenko (1989), which states that any continuous function on a compact domain can be approximated with any precision by an appropriate neural network with sufficient width and depth. Beyond the classical result stated above, several types of variants of the universal approximation theorem have also been investigated under different conditions.

AAmong a wide variety of deep neural networks, convolutional neural networks (CNNs) have achieved impressive performance for real applications. In particular, almost all of state-of-the-art models for image recognition are based on CNNs. These successes are closely related to the property that performing CNNs commute with translation on pixel coordinate. That is, CNNs can conserve symmetry about translation in image data. In general, this kind of property for symmetry is known as the *equivariance*, which is a generalization of the *invariance*. When a data distribution has some symmetry and the task to be solved relates to the symmetry, data processing is desired to be equivariant on the symmetry. In recent years, different types of symmetry have been focused per each task, and it has been proven that CNNs can approximate arbitrary equivariant data processing for specific symmetry. These results are mathematically captured as the universal approximation for equivariant maps and represent the theoretical validity of the use of CNNs.

In order to theoretically correctly handle symmetric structures, we have to carefully consider the structure of data space where data distributions are defined. For example, in image recognition tasks, image data are often supposed to have symmetry for translation. When each image data is acquired, there are finite pixels equipped with an image sensor, and an image data is represented by a finite-dimensional vector in a Euclidean space $\mathbb{R}^d$, where d is the number of pixels. However, we note that the finiteness of pixels stems from the limit of the image sensor and raw scenes behind image data x are thought to be modelled by elements $\tilde{x}$ in $\mathbb{R}^{\mathcal{S}}$ with continuous space coordinates $\mathcal{S}$, where $\mathbb{R}^{\mathcal{S}}$ is a set of map from $\mathcal{S}$ to $\mathbb{R}$. Then, $\tilde{x}$ is regarded as a functional representation of data x. To appropriately formulate data symmetry, we treat both typical data representation in finite-dimensional settings and functional representation in infinite-dimensional settings in a unified manner.

1.1 Related Works and Our Contributions

Functional Representation and Symmetry. Gordon et al. (2019) point out that, although data symmetry cannot be treated in the set of finite pixels, it can be represented by a group action on continuous space coordinates $\mathcal{S}$. Moreover, the authors also show that functional representations $\tilde{x}$ instead of image data x are suitable for analyzing and considering data processing compatible with data symmetry. As a related study about symmetry-preserving processing, Finzi et al. (2020) propose group convolution of functional representations and investigate practical computational methods such as discretization and localization.

Universal approximation. The universal approximation theorem, which is the main objective of this paper, is one of the most classical mathematical theorems of neural networks. The universal approximation theorem states that a feedforward fully-connected network (FNN) with a single hidden layer containing finite neurons can approximate a continuous function on a compact subset of $\mathbb{R}^d$. Cybenko (1989) proved this theorem for the sigmoid activation function. After his work, some researchers showed similar results to generalize the sigmoidal function to a larger class of activation functions as Barron (1994), Hornik et al. (1989), Funahashi (1989), Kůrková (1992) and Sonoda & Murata (2017). These results were approximations to functional representations between finite-dimensional vector spaces, but recently Guss & Salakhutdinov (2019) generalized them to continuous maps between infinite-dimensional function spaces in Guss & Salakhutdinov (2019).

Group invariant/equivariant networks. Symmetry is mathematically described in terms of groups and has become an essential concept in machine learning. Theoretically speaking, the concept of group representation (symmetric) networks was introduced by Wood & Shawe-Taylor (1996). In machine learning, deep symmetry networks are designed in Gens & Domingos (2014) as a generalization of convnets that form feature maps on top of arbitrary symmetry groups. The models of the permutation-invariant/equivariant tasks are designed in Zaheer et al. (2017) to obtain excellent results in population statistic estimation, point group classification, set expansion, and anomaly detection. Maron et al. (2019a; 2020) consider and develop a theory of invariant tensor networks for finite groups.

Universal approximation of group invariant/equivariant networks. Yarotsky (2018) focused on the action of compact groups and provided the network's architectures based on polynomial layers. He also describes a convnet-like model that is a universal approximator of equivariant transformations for the group SE(2) of rigid two-dimensional euclidean motions. Maron et al. (2019b) considered a G-invariant model for an arbitrary subrepresentation G of a symmetry group, using an invariant tensor network. Sannai et al. (2019); Ravanbakhsh (2020) considered a G-equivariant model and proved universal approximation property of them by attributing it to the results of Maron et al. (2019b). Thereby, they proved *universal approximation theorem of invariant/equivariant networks* for compact and finite group actions.

Group convolutional neural networks. Convolutional neural networks (CNNs) have a built-in understanding of the local properties and translational symmetry inherent in many learning problems. Motivated by CNN's experimental success, many researchers have generalized this by using group theory. Cohen et al. (2019) considered group convolution on a homogeneous space and proved that a linear equivariant map is always convolution-like. Kondor & Trivedi (2018) proved that the necessary and sufficient condition for a neural net with a constraint structure associated with a homogeneous space to become an equivariant map is that it becomes a group CNN. Petersen & Voigtlaender (2020) showed that a group CNN, which approximates the equivariant map for a finite group G, corresponds to a fully connected neural net, which approximates the projection of the map to the first component of the map and gives the same approximation accuracy.

Our contributions.

- We clarify some of the basic properties of equivariant maps. In particular, we show that equivariant maps' generators play an essential role in the analysis of equivariant maps.
- We formulate neural networks in general settings. Because of this formulation, we can handle finite- and infinite-dimensional settings in a unified manner.
- We provide a theorem that can convert FNNs to CNNs. Using this theorem, we can induce the proofs of universal approximation theorems for equivariant maps by CNN to those

for continuous maps by FNNs. In particular, we obtain the first universal approximation theorems for equivariant maps in infinite-dimensional settings.

- We also provide a novel universal approximation theorem for Lipschitz maps by FNNs in an infinite-dimensional setting. This result enables us to derive universal approximation theorems for equivariant maps for non-compact groups while almost all of the existing works could not handle the action of non-compact groups.

2 Group Equivariance

2.1 Preliminaries

We introduce definitions and terminology used in the later discussion.

Functional representation. In this paper, sets denoted by $\mathcal{S}$, $\mathcal{T}$ and G are assumed to be locally compact, σ-compact, Hausdorff spaces. When $\mathcal{S}$ is a set, we denote by $\mathbb{R}^{\mathcal{S}}$ the set of all maps from $\mathcal{S}$ to $\mathbb{R}$ and by $\|\cdot\|_\infty$ the supremum norm. We call $\mathcal{S}$ of $\mathbb{R}^{\mathcal{S}}$ the index set. We denote by $\mathcal{C}(\mathcal{S})$ the set of all continuous maps from $\mathcal{S}$ to $\mathbb{R}$. We denote by $\mathcal{C}_0(\mathcal{S})$ the set of continuous functions from $\mathcal{S}$ to $\mathbb{R}$ which vanish at infinity[1]. For a Borel space $\mathcal{S}$ with some measure μ, we denote the set of L^p-integrable functions from $\mathcal{S}$ to $\mathbb{R}$ with respect to μ as $L^p(\mathcal{S})$. For a subset $\mathcal{B} \subset \mathcal{S}$, the restriction map $R_{\mathcal{B}} : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{B}}$ is defined by $R_{\mathcal{B}}(x) = x|_{\mathcal{B}}$, where $x \in \mathbb{R}^{\mathcal{S}}$ and $x|_{\mathcal{B}}$ is the restriction of the domain of x onto $\mathcal{B}$. The indicator function on $\mathcal{B} \subset \mathcal{S}$ is denoted by $\mathbf{1}_{\mathcal{B}}$.

When $\mathcal{S}$ is a finite set, $\mathbb{R}^{\mathcal{S}}$ is identified with the finite-dimensional Euclidean space $\mathbb{R}^{|\mathcal{S}|}$, where $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. In this sense, $\mathbb{R}^{\mathcal{S}}$ for general sets $\mathcal{S}$ is a generalization of Euclidean spaces. However, $\mathbb{R}^{\mathcal{S}}$ itself is often intractable. In such cases, we instead consider $\mathcal{C}(\mathcal{S})$, $\mathcal{C}_0(\mathcal{S})$ or $L^p(\mathcal{S})$ as relatively tractable subspaces of $\mathbb{R}^{\mathcal{S}}$.

Group action. We denote the identity element in a group G by 1. We assume that the action of a group G on a set $\mathcal{S}$ is continuous. We denote by $g \cdot s$ the left action of $g \in G$ to $s \in \mathcal{S}$. Then we call $G_s := \{g \cdot s | g \in G\}$ the orbit of $s \in \mathcal{S}$. From the definition, we have $\mathcal{S} = \bigcup_{s \in \mathcal{S}} G_s$. When a subset $\mathcal{B} \subset \mathcal{S}$ is the set of representative elements from all orbits, it satisfies the disjoint condition $\mathcal{S} = \bigsqcup_{s \in \mathcal{B}} G_s$. Then, we call $\mathcal{B}$ a base space[2] and define the projection $P_{\mathcal{B}} : \mathcal{S} \to \mathcal{B}$ by mapping $s \in \mathcal{S}$ to the representative element in $\mathcal{B} \cap G_s$. When a group G acts on sets $\mathcal{S}$ and $\mathcal{T}$, the action of G on the product space $\mathcal{S} \times \mathcal{T}$ is defined by $g \cdot (s, t) := (g \cdot s, g \cdot t)$. When a group G acts on a index set $\mathcal{S}$, the G-translation operators $T_g : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{S}}$ for $g \in G$ are defined by $T_g[x](s) := x(g^{-1} \cdot s)$, where $x \in \mathbb{R}^{\mathcal{S}}$ and $s \in \mathcal{S}$. We often denote $T_g[x]$ simply by $g \cdot x$ for brevity. Then, group translation determine the action of G on $\mathbb{R}^{\mathcal{S}}$[3].

2.2 Group Equivariant Maps

In this section, we introduce group equivariant maps and show their basic properties. First, we define group equivariance.

Definition 1 (Group Equivariance)**.** *Suppose that a group G acts on sets $\mathcal{S}$ and $\mathcal{T}$. Then, a map $F : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$ is called G-equivariant when $F(g \cdot x) = g \cdot F(x)$ holds for any $g \in G$ and $x \in \mathbb{R}^{\mathcal{S}}$.*

In order to clarify the degree of freedom of equivariant maps, we define the generator of equivariant maps.

Definition 2 (Generator)**.** *Let $\mathcal{B} \subset \mathcal{T}$ be a base space with respect to the action of G on $\mathcal{T}$. For a G-equivariant map $F : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$, we call $F_{\mathcal{B}} := R_{\mathcal{B}} \circ F$ the generator of F.*

The following theorem shows that equivariant maps can be represented by their generators.

Theorem 3 (Degree of Freedom of Equivariant Maps)**.** *Let a group G act on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B} \subset \mathcal{T}$ a base space. Then, a G-equivariant map $F : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$ can be represented using its*

[1] A function f on a locally compact space is said to vanish at infinity if, for any ϵ, there exists a compact subset $\mathcal{K} \subset \mathcal{S}$ such that $\sup_{s \in \mathcal{S} \setminus \mathcal{K}} |f(s)| < \epsilon$.

[2] The choice of the base space is not unique in general. However, the topological structure of a base space can be induced by the quotient space $\mathcal{S}/G$.

[3] We note that $T_g \circ T_{g'} = T_{g'g}$ and the group translation operator is the action of G on $\mathbb{R}^{\mathcal{S}}$ from the right.

generator as

$$F[x](t) = F_{\mathcal{B}}[g_t^{-1} \cdot x](P_{\mathcal{B}}(t)), \tag{1}$$

where $g_t \in G$ is an arbitrary element which satisfies $g_t \cdot P_{\mathcal{B}}(t) = t$. Conversely, for an arbitrary map $F_{\mathcal{B}} : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{B})$, a map F defined by (1) is an equivariant map whose generator equals $F_{\mathcal{B}}$.

Theorem 3 clarifies the rigidity and flexibility of the class of equivariant maps. That is, equivariant maps are completely rigid given generators in the sense that the generator determines those. On the other hand, the generators of equivariant maps are entirely flexible because they have no restrictions on constructing equivariant maps.

From the following proposition, the distance between equivariant maps is calculated from their generators.

Proposition 4 (Isometric Restriction)**.** *Let a group G act on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B} \subset \mathcal{T}$ an arbitrary base space. The restriction $R_{\mathcal{B}}$ onto $\mathcal{B}$ is isometry from equivariant maps. That is, for two G-equivariant maps F and $\tilde{F} : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$,*

$$\|F - \tilde{F}\|_\infty = \|F_{\mathcal{B}} - \tilde{F}_{\mathcal{B}}\|_\infty. \tag{2}$$

We immediately obtain the following corollary from Proposition 4.

Corollary 5 (Identity Condition)**.** *Let a group G act on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B} \subset \mathcal{T}$ an arbitrary base space. Let F and $\tilde{F} : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$ be G-equivariant maps. Then, $F = \tilde{F}$ if and only if $F_{\mathcal{B}} = \tilde{F}_{\mathcal{B}}$.*

3 Fully-connected and Group Convolutional Neural Networks

3.1 Fully-connected Neural Networks

In order to define neural networks, we introduce some notions. A map $A : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$ is called a bounded affine map if there exist a bounded linear map $W : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$ and an element $b \in \mathbb{R}^{\mathcal{T}}$ such that

$$A[x] = W[x] + b. \tag{3}$$

Guss & Salakhutdinov (2019) provide the following lemma, which is useful to handle bounded affine maps.

Lemma 6 (Integral Form, Guss & Salakhutdinov (2019))**.** *Suppose that $\mathcal{S}$ and $\mathcal{T}$ are locally compact, σ-compact, Hausdorff, measurable spaces. For a bounded linear map $W : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{T})$, there exist a Borel regular measure μ on $\mathcal{S}$ and a weak* continuous family of functions $\{w(t, \cdot)\}_{t \in \mathcal{T}} \subset L^1(\mu)$ on $\mathcal{S}$ such that the following holds for any $x \in \mathcal{C}(\mathcal{S})$:*

$$W[x](t) = \int_{\mathcal{S}} w(t, s) x(s) d\mu(s). \tag{4}$$

To use the integral form, we assume in the following that the input and output spaces of A are the class of continuous maps $\mathcal{C}(\mathcal{S})$ and $\mathcal{C}(\mathcal{T})$ instead of $\mathbb{R}^{\mathcal{S}}$ and $\mathbb{R}^{\mathcal{T}}$, respectively. Using the integral form, a bounded affine map A is represented by

$$A_{\mu,w,b}[x](t) = \int_{\mathcal{S}} w(t, s) x(s) d\mu(s) + b(t).$$

In particular, we call the above μ the measure of $A = A_{\mu,w,b}$.

A continuous function $\rho : \mathbb{R} \to \mathbb{R}$ induces the activation map $\alpha_\rho : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{S})$ which is defined by $\alpha_\rho(x) := \rho \circ x \in \mathcal{C}(\mathcal{S})$ for $x \in \mathcal{C}(\mathcal{S})$. However, for brevity, we denote α_ρ by ρ. Then, we can define fully-connected neural networks in general settings.

Definition 7 (Fully-connected Neural Networks)**.** *Let $L \in \mathbb{N}$. A fully-connected neural network with L layers is a composition map of bounded affine maps $(A_1, \ldots, A_L)$ and an activation map ρ represented by*

$$\phi := A_L \circ \rho \circ A_{L-1} \circ \cdots \circ \rho \circ A_1, \tag{5}$$

where $A_\ell : \mathcal{C}(\mathcal{S}_{\ell-1}) \to \mathcal{C}(\mathcal{S}_\ell)$ are affine maps for some sequence of sets $\{\mathcal{S}_\ell\}_{\ell=0}^L$. Then, we denote by $\mathcal{N}_{\mathsf{FNN}}(\rho, L; \mathcal{S}_0, \mathcal{S}_L)$ the set of all fully-connected neural networks from $\mathcal{C}(\mathcal{S}_0)$ to $\mathcal{C}(\mathcal{S}_L)$ with L layers and an activation function ρ.

We denote the measure of the affine map A_1 in the first layer of a fully-connected neural network ϕ by μ_ϕ. This measure μ_ϕ is used to describe a condition in the main theorem (Theorem 11).

3.2 Group Convolutional Neural Networks

Definition 8 (Group Convolution)**.** *Suppose that a group G acts on sets $\mathcal{S}$ and $\mathcal{T}$. For a G-invariant measure ν on $\mathcal{S}$, G-invariant functions $v : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ and $b \in \mathcal{C}(\mathcal{T})$, the biased G-convolution $C_{\nu,v,b} : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{T})$ is defined as*

$$C_{\nu,v,b}[x](t) := \int_{\mathcal{S}} v(t,s)x(s)d\nu(s) + b(t). \tag{6}$$

In the right hand side, we call the first term the G-convolution and the second term the bias term.

In the following, we denote $C_{\nu,v,b}$ by C for brevity.

Definition 8 includes existing definitions of group convolution as follows. When $\mathcal{S} = \mathcal{T} = G$, the group G acts on $\mathcal{S}$ and $\mathcal{T}$ by left translations. Then, (6) without the bias term (i.e., $b = 0$) is described as

$$C[x](g) = \int_{\mathcal{G}} v(g,h)x(h)d\nu(h) = \int_{\mathcal{G}} \tilde{v}(h^{-1}g)x(h)d\nu(h),$$

where $\tilde{v}(g) := v(g,1)$[4]. This is a popular definition of group convolution between two functions on G. Further, when $\mathcal{S} = G \times \mathcal{B}$ and $\mathcal{T} = G \times \mathcal{B}'$, (6) without the bias term is described as

$$C[x](g,t) = \int_{\mathcal{G}\times\mathcal{B}} v((g,\tau),(h,\varsigma))x(h,\varsigma)d\nu(h,\varsigma) = \int_{\mathcal{G}\times\mathcal{B}} \tilde{v}(h^{-1}g,\tau,\varsigma)x(h,\varsigma)d\nu(h,\varsigma),$$

where $\tilde{v}(g,\tau,\varsigma) := v((g,\tau),(1,\varsigma))$. This coincides with the definition of group convolution in Finzi et al. (2020). We note that Finzi et al. (2020) also proposes discretization and localization of the above group convolution for practical implementation.

In conventional convolution used for image recognition, G represents spatial information such as pixel coordinate, $\mathcal{B}$ and $\mathcal{B}'$ correspond to channels in consecutive layers ℓ and $\ell + 1$ respectively, and v corresponds to a filter. In applications, the filter v is expected to have compact support or be short-tailed on G as in a 3×3 convolution filter in discrete convolution. In particular, when v is allowed to be the Dirac delta or highly peaked around a single point in G, such convolution can be interpreted as the 1×1 convolution.

Then, we define group convolutional neural networks as follows.

Definition 9 (Group Convolutional Neural Networks)**.** *Let $L \in \mathbb{N}$. A G-convolutional neural network with L layers is a composition map of biased convolutions $C_\ell : \mathcal{C}(\mathcal{S}_{\ell-1}) \to \mathcal{C}(\mathcal{S}_\ell)$ ($\ell = 1, \ldots, L$) for some sequence of spaces $\{\mathcal{B}_\ell\}_{\ell=0}^{L}$ and an activation map with ρ as*

$$\Phi := C_L \circ \rho \circ C_{L-1} \circ \cdots \circ \rho \circ C_1. \tag{7}$$

Then, we denote by $\mathcal{N}_{\mathsf{CNN}}(G, \rho, L; \mathcal{S}_0, \mathcal{S}_L)$ the set of all G-convolutional neural networks from $\mathcal{C}(\mathcal{S}_0)$ to $\mathcal{C}(\mathcal{S}_L)$ with respect to a group G with L layers and a fixed activation function ρ.

We easily verify the following proposition.

Proposition 10. *A G-convolutional neural network is G-equivariant.*

In particular, each biased G-convolution $C_{\nu,v,b}$ is G-equivariant. Conversely, Cohen et al. (2019) showed that a G-equivariant *linear* map is represented by some G-convolution without the bias term when G is locally compact and unimodular, and the action of a group is transitive (i.e., $\mathcal{B}$ consists of only a single element).

[4] A bivariate G-invariant function $v : G \times G \to \mathbb{R}$ is determined by the univariate function $\tilde{v} : G \to \mathbb{R}$ because $v(g,h) = v(h^{-1}g, h^{-1}h) = v(h^{-1}g, 1) = \tilde{v}(h^{-1}g)$.

3.3 CONVERSION THEOREM

In this section, we introduce the main theorem (Theorem 11), which is an essential part of obtaining universal approximation theorems for equivariant maps by group CNNs.

Theorem 11 (Conversion Theorem)**.** *Suppose that the action of a group G on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B}_\mathcal{S} \subset \mathcal{S}$ and $\mathcal{B}_\mathcal{T} \subset \mathcal{T}$ are base spaces. We assume the following condition:*

(C1) there exist subgroups $H_\mathcal{T} \subset H_\mathcal{S} \subset G$ such that $\mathcal{S} = G/H_\mathcal{S} \times \mathcal{B}_\mathcal{S}$ and $\mathcal{T} = G/H_\mathcal{T} \times \mathcal{B}_\mathcal{T}$.

Further, suppose an FNN $\phi : E \to \mathcal{C}_0(\mathcal{B}_\mathcal{T})$ with a Lipschitz activation function satisfies

(C2) there exists a G-invariant locally finite measure ν on $\mathcal{S}$ such that $\mu_\phi \ll \nu$[5].

Then, for any $\epsilon > 0$ and compact subset $E \subset \mathcal{C}_0(\mathcal{S})$, there exists a CNN $\Phi : E \to \mathcal{C}_0(\mathcal{T})$ such that the number of layers of Φ equals that of ϕ and

$$\|R_{\mathcal{B}_\mathcal{T}} \circ \Phi - \phi\|_\infty \leq \epsilon. \tag{8}$$

Moreover, for any G-equivariant map $F : \mathcal{C}_0(\mathcal{S}) \to \mathcal{C}_0(\mathcal{T})$, the following holds:

$$\|F|_E - \Phi\|_\infty \leq \|F_{\mathcal{B}_\mathcal{T}}|_E - \phi\|_\infty + \epsilon. \tag{9}$$

We provide the proof of Theorem 11 in Section B. In Theorem 11, we note that a group G may be non-compact or non-commutative. Further, both discrete (especially finite) and continuous groups are allowed.

Application to Universal Approximation Theorems. The conversion theorem can be used to convert a universal approximation theorem by FNNs to a universal approximation theorem for equivariant maps by CNNs as follows. Suppose that the existence of an FNN ϕ which satisfies $\|F_\mathcal{B}|_E - \phi\|_\infty \leq \epsilon$ using some universal approximation theorem by FNNs. Then, Theorem 11 guarantees the existence of a CNN Φ which satisfies $\|F|_E - \Phi\|_\infty \leq 2\epsilon$. In other words, if an FNN can approximate the generator of the target equivariant map on E, then there exists a CNN which approximates the whole of the equivariant map on E.

In the following, we consider the conditions (C1) and (C2).

In (C1), the subgroup $H_\mathcal{S} \subset G$ (resp. $H_\mathcal{T}$) represents the stabilizer group of the action of G on $\mathcal{S}$ (resp. $\mathcal{T}$). Thus, (C1) requires that the stabilizer group on every point in $\mathcal{S}$ (resp. $\mathcal{T}$) is isomorphic to the common subgroup $H_\mathcal{S}$ (resp. $H_\mathcal{T}$). When the group action satisfies some moderate conditions, such a requirement is known to be satisfied for *most* points in the set. As a theoretical result, the principal orbit type theorem (cf. Theorem 1.32, Meinrenken (2003)) guarantees that, if the group action on a manifold $\mathcal{S}$ is proper and $\mathcal{S}/G$ is connected, there exist a dense subset $\mathcal{S}' \subset \mathcal{S}$ and a subgroup $H_\mathcal{S} \subset G$ called a principal stabilizer such that the stabilizer group on every point in $\mathcal{S}'$ is isomorphic to $H_\mathcal{S}$.

Further, (C1) assumes that the sets $\mathcal{S}$ and $\mathcal{T}$ have the direct product form of some quotient group G/H and a base space $\mathcal{B}$. In many practical cases, a set $\mathcal{S}$ on which G acts can be regarded as such a direct product form. For example, when the action is transitive, the direct product decomposition trivially holds with the base space that consists of a single point. Even when the set $\mathcal{S}$ itself is not rigorously represented by the direct product form, removing some *small* subset $\mathcal{N} \subset \mathcal{S}$, the complement $\mathcal{S} \setminus \mathcal{N}$ can be often represented by the direct form. For example, when $G = SO(d)$ acts on the set $\mathcal{S} = \mathbb{R}^d$ as rotation around the origin $\mathcal{N} = \{0\}$, the stabilizer of every point in $\mathbb{R}^d \setminus \{0\}$ is known to be isomorphic to $H = SO(d-1)$. Then, $\mathbb{R}^d \setminus \{0\}$ can be represented by the direct product form $SO(d)/SO(d-1) \times \mathbb{R}_+$ with the base space $\mathcal{B} = \mathbb{R}_+$. In applications, removing only the small subset $\mathcal{N}$ is expected to be negligible.

We mention the relation to the invariant case. When the stabilizer in $\mathcal{T}$ satisfies $H_\mathcal{T} = G$, the group action is said to be G-invariant. However, because of the condition $H_\mathcal{T} \subset H_\mathcal{S}$ in (C1), Theorem 11 can not virtually apply to the invariant case. This restriction is similar to existing studies, where the invariant case is separately handled from the equivariant case (Keriven & Peyré (2019); Maehara & NT (2019); Sannai et al. (2019)).

[5] $\mu_\phi \ll \nu$ means that μ_ϕ is absolutely continuous with respect to ν.

Next, we provide some remarks on the condition (C2). Let us consider two representative settings of a set $\mathcal{S}$. The first case is the setting where $\mathcal{S}$ is finite. When a G-invariant measure ν has a positive value on every singleton in $\mathcal{S}$, ν satisfies (C2) for an arbitrary measure μ_ϕ on $\mathcal{S}$. In particular, the counting measure on $\mathcal{S}$ is invariant and satisfies (C2). The second case is the setting where $\mathcal{S}$ is a Euclidean space $\mathbb{R}^d$, and μ_ϕ is the Lebesgue measure. Then, (C2) is satisfied with invariant measures on the Euclidean space for various group actions, including translation, rotation, scaling, and a Euclidean group.

Lastly, we give a general method to construct ν in (C2) for a compact-group action. When μ_ϕ is locally finite and continuous with respect to the action of a compact group G[6], the measure $\nu := \nu_G * \mu_\phi$ on $\mathcal{S}$ for a Haar measure ν_G on G satisfies (C2), where $(\nu_G * \mu_\phi)(A) := \int_G \mu_\phi(g^{-1} \cdot A) d\nu_G(g)$.

4 Universal Approximation Theorems for Equivariant Maps

4.1 Universal Approximation Theorem in Finite Dimension

We review the universal approximation theorem in finite-dimensional settings. Here, we regard the finite-dimensional Euclidean space $\mathbb{R}^d$ as $\mathcal{C}([d])$ and d-dimensional vectors in $\mathbb{R}^d$ as maps from $[d]$ to $\mathbb{R}$, where $[d] := \{1, 2, \ldots, d\}$ for $d \in \mathbb{N}$. Then, an affine map $\phi : \mathbb{R}^d \to \mathbb{R}^{d'}$ is parameterized by a weight matrix $W = [w(t,s)]_{s\in[d], t\in[d']} : \mathbb{R}^d \to \mathbb{R}^{d'}$ and a bias vector $b = [b(t)]_{t\in[d']} \in \mathbb{R}^{d'}$, and (4) induces the following form which is often used in the literature on neural networks:

$$\phi[x](t) = \sum_{s=1}^{d} x(x) w(t,s) + b(t). \tag{10}$$

Cybenko (1989) derived the following seminal results about universal approximation theorem in finite-dimensional settings.

Theorem 12 (Universal Approximation for Continuous Maps by FNNs, Cybenko (1989))**.** *Let an activation function $\rho : \mathbb{R} \to \mathbb{R}$ be non-constant, bounded and continuous. Let $F : \mathbb{R}^d \to \mathbb{R}^{d'}$ be a continuous map. Then, for any compact $E \subset \mathbb{R}^d$ and $\epsilon > 0$, there exists a two-layer fully connected neural network $\phi_E \in \mathcal{N}_{\mathsf{FNN}}(\rho, 2; [d], [d'])$ such that $\|F|_E - \phi_E\|_\infty < \epsilon$.*

Since $\mathcal{C}_0(\mathcal{U}) = \mathbb{R}^{|\mathcal{U}|}$ for a finite set $\mathcal{U}$, we obtain the following theorem by combining Theorem 11 with Theorem 12.

Theorem 13 (Universal Approximation for Equivariant Continuous Maps by CNNs)**.** *Let an activation function $\rho : \mathbb{R} \to \mathbb{R}$ be non-constant, bounded and Lipschitz continuous. Suppose that a finite group G acts on finite sets $\mathcal{S}$ and $\mathcal{T}$ and (C1) in Thoerem 11 holds. Let $F : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{T}|}$ be a G-equivariant continuous map. For any compact set $E \subset \mathbb{R}^{|\mathcal{S}|}$ and $\epsilon > 0$, there exists a two-layer convolutional neural network $\Phi_E \in \mathcal{N}_{\mathsf{CNN}}(\rho, 2; |\mathcal{S}|, |\mathcal{T}|)$ such that $\|F|_E - \Phi_E\|_\infty < \epsilon$.*

We note that Petersen & Voigtlaender (2020) obtained a similar result to Theorem 13 in the case of finite sets.

4.2 Universal Approximation Theorem in Infinite Dimension

Recently, Guss & Salakhutdinov (2019) derived the following theorem in infinite-dimensional settings.

Theorem 14 (Universal Approximation for Continuous Maps by FNNs, Guss & Salakhutdinov (2019))**.** *Let an activation function $\rho : \mathbb{R} \to \mathbb{R}$ be continuous and non-polynomial. Let $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}^{d'}$ be compact domains. Let $F : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{T})$ be a continuous map. Then, for any compact $E \subset \mathcal{C}(\mathcal{S})$ and $\epsilon > 0$, there exist $N \in \mathbb{N}$ and a two-layer fully connected neural network $\phi_E = A_2 \circ \rho \circ A_1 \in \mathcal{N}_{\mathsf{FNN}}(\rho, 2; \mathcal{S}, \mathcal{T})$ such that $A_1[\cdot] = W^{(1)}[\cdot] + b^{(1)} : E \to \mathcal{C}([N]) = \mathbb{R}^N$, $A_2[\cdot] = W^{(2)}[\cdot] + b^{(2)} : \mathbb{R}^N \to \mathcal{C}(\mathcal{T})$, μ_{ϕ_E} is the Lebesgue measure, and $\|F|_E - \phi_E\|_\infty < \epsilon$.*

[6] A measure μ_ϕ is said to be continuous with respect to the action of a group G if $\mu_\phi(g \cdot A)$ is continuous with respect to $g \in G$ for all Borel set $A \subset \mathcal{S}$.

Thus, any continuous function can be approximated by an FNN whose neurons in the hidden layer is finite. Note that it is assumed in Theorem 14 that the index set $\mathcal{S}$ in the input layer and $\mathcal{T}$ in the output layer are compact. Thus, combining the conversion theorem with Theorem 14, we can derive a corresponding universal approximation theorem for equivariant maps with respect to compact groups. However, the compactness condition for $\mathcal{S}$ and $\mathcal{T}$ is a crucial shortcoming to handle the action of non-compact groups such as translation or scaling.

In order to overcome the above obstacle, we replace the space of continuous functions on compact sets by $\mathcal{C}_0(\mathbb{R}^d)$. Then, we can show a novel universal approximation theorem for Lipschitz maps by FNNs as follows.

Theorem 15 (Universal Approximation for Lipschitz Maps by FNNs)**.** *Let an activation function $\rho : \mathbb{R} \to \mathbb{R}$ be continuous and non-polynomial. Let $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}^{d'}$ be domains. Let $F : \mathcal{C}_0(\mathcal{S}) \to \mathcal{C}_0(\mathcal{T})$ be a Lipschitz map. Then, for any compact $E \subset \mathcal{C}_0(\mathcal{S})$ and $\epsilon > 0$, there exist $N \in \mathbb{N}$ and a two-layer fully connected neural network $\phi_E = A_2 \circ \rho \circ A_1 \in \mathcal{N}_{\mathsf{FNN}}(\rho, 2; \mathcal{S}, \mathcal{T})$ such that $A_1[\cdot] = W^{(1)}[\cdot] + b^{(1)} : E \to \mathcal{C}([N]) = \mathbb{R}^N$, $A_2[\cdot] = W^{(2)}[\cdot] + b^{(2)} : \mathbb{R}^N \to \mathcal{C}(\mathcal{T})$, μ_{ϕ_E} is the Lebesgue measure, and $\|F|_E - \phi_E\|_\infty < \epsilon$.*

We note that $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}^{d'}$ in Theorem 15 are allowed to be non-compact unlike Theorem 14. We provide proof of Theorem 15 in the appendix. Combining Theorem 11 with Theorem 15, we obtain the following theorem.

Theorem 16 (Universal Approximation for Equivariant Lipschitz Maps by CNNs)**.** *Let an activation function $\rho : \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous and non-polynomial. Suppose that a group G acts on $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}^{d'}$, and (C1) and (C2) in Thoerem 11 hold for the Lebesgue measure μ_ϕ. Let $F : \mathcal{C}_0(\mathcal{S}) \to \mathcal{C}_0(\mathcal{T})$ be a G-equivariant Lipschitz map. Then, for any compact set $E \subset \mathcal{C}_0(\mathcal{S})$ and $\epsilon > 0$, there exists a two-layer convolutional neural network $\Phi_E \in \mathcal{N}_{\mathsf{CNN}}(\rho, 2; \mathcal{S}, \mathcal{T})$ such that $\|F|_E - \Phi_E\|_\infty < \epsilon$.*

In Theorem 16, we assumed sets $\mathcal{S}$ and $\mathcal{T}$ to be subspaces of Euclidean spaces. However, in the conversion theorem (Theorem 11), sets $\mathcal{S}$ and $\mathcal{T}$ do not need to be subspaces of Euclidean spaces and may have a more general topological structure. Thus, if there is a universal approximation theorem in non-Euclidean spaces such as graphs, manifolds, or more general topological spaces (Courrieu (2005); Kratsios (2019)), we may be able to combine it with the conversion theorem and derive its equivariant version. Related to this, investigating the connection to equivariant graph neural networks (Keriven & Peyré (2019); Maehara & NT (2019)) is an interesting future work.

5 Conclusion

We have considered universal approximation theorems for equivariant maps by group CNNs. In order to prove the theorems, we first clarified several general properties of equivariant maps. One of the fundamental properties is that any equivariant map has a one-to-one correspondence to its generator. In particular, we showed that an equivariant map is uniquely constructed from the generator by utilizing the algebraic nature of the equivariance. Thus, if the generator of an equivariant map can be approximated by some map, we can also construct an approximator of the whole of the equivariant map. Moreover, we showed that, when we can take a fully-connected neural network to approximate the generator, the approximator of the equivariant map constructed above can be described as a group CNN. In this way, the universal approximation for equivariant maps by group CNNs can be obtained through the universal approximation for the generator by FNNs. We have described FNNs and group CNNs in an abstract way. In particular, theoretical results in this paper can apply to not only finite-dimensional settings but also infinite-dimensional settings. We provided a novel universal approximation theorem by FNNs in the infinite dimension, where the support of the input functions is unbounded. Using this result, we obtained the universal approximation theorem for equivariant maps for non-compact groups, which could not be handled in most existing studies.

Lastly, we mention the problem of computational complexity. Although group convolution in each layer of CNNs can be implemented by, e.g., discretization and localization as in Finzi et al. (2020), such implementation cannot be applied to high-dimensional groups without a significant reduction in computational complexity. In order to use group CNNs for actual machine-learning problems, it is required to construct effective architecture for practical implementation.

REFERENCES

Andrew R Barron. Approximation and estimation bounds for artificial neural networks. *Machine learning*, 14(1):115–133, 1994.

Taco S Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. In *Advances in Neural Information Processing Systems*, pp. 9142–9153, 2019.

Pierre Courrieu. Function approximation on non-euclidean spaces. *Neural Networks*, 18(1):91–102, 2005.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. *arXiv preprint arXiv:2002.12880*, 2020.

Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural networks*, 2(3):183–192, 1989.

Robert Gens and Pedro M Domingos. Deep symmetry networks. In *Advances in Neural Information Processing Systems*, pp. 2537–2545, 2014.

Jonathan Gordon, Wessel P Bruinsma, Andrew YK Foong, James Requeima, Yann Dubois, and Richard E Turner. Convolutional conditional neural processes. *arXiv preprint arXiv:1910.13556*, 2019.

William H Guss and Ruslan Salakhutdinov. On universal approximation by neural networks with uniform guarantees on approximation of infinite dimensional maps. *arXiv preprint arXiv:1910.01545*, 2019.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. In *Advances in Neural Information Processing Systems*, pp. 7092–7101, 2019.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. *arXiv preprint arXiv:1802.03690*, 2018.

Anastasis Kratsios. The universal approximation property: Characterizations, existence, and a canonical topology for deep-learning. *arXiv preprint arXiv:1910.03344*, 2019.

Mateusz Krukowski. Frechet-kolmogorov-riesz-weil's theorem on locally compact groups via arzela-ascoli's theorem. *arXiv preprint arXiv:1801.01898*, 2018.

Vvera Kůrková. Kolmogorov' s theorem and multilayer neural networks. *Neural networks*, 5(3): 501–506, 1992.

Takanori Maehara and Hoang NT. A simple proof of the universality of invariant/equivariant graph neural networks. *arXiv preprint arXiv:1910.03802*, 2019.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *International Conference on Learning Representations*, 2019a. URL `https://openreview.net/forum?id=Syx72jC9tm`.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. *Proceedings of the 36th International Conference on Machine Learning*, 97, 2019b.

Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. On learning sets of symmetric elements. *arXiv preprint arXiv:2002.08599*, 2020.

Eckhard Meinrenken. Group actions on manifolds. *Lecture Notes, University of Toronto, Spring*, 2003, 2003.

Philipp Petersen and Felix Voigtlaender. Equivalence of approximation by convolutional neural networks and fully-connected networks. *Proceedings of the American Mathematical Society*, 148 (4):1567–1581, 2020.

Siamak Ravanbakhsh. Universal equivariant multilayer perceptrons. *arXiv preprint arXiv:2002.02912*, 2020.

Akiyoshi Sannai, Yuuki Takai, and Matthieu Cordonnier. Universal approximations of permutation invariant/equivariant functions by deep neural networks. *arXiv preprint arXiv:1903.01939*, 2019.

Sho Sonoda and Noboru Murata. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2):233–268, 2017.

Jeffrey Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete applied mathematics*, 69(1-2):33–60, 1996.

Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *arXiv preprint arXiv:1804.10306*, 2018. URL: `https://arxiv.org/abs/1804.10306`.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in neural information processing systems*, pp. 3391–3401, 2017.