Neural Networks for Learning Counterfactual G-Invariances from Single Environments

"Fixing the Image Rotation Problem"

J. Setpal

September 26, 2024

Outline

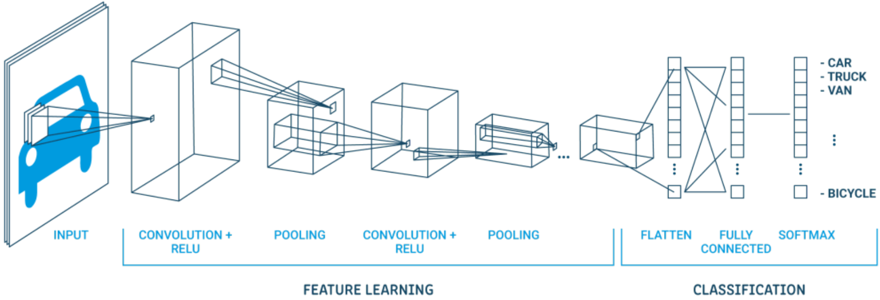
1. Motivation
2. Set Theory
3. Leveraging Set Theory (Fun Part)

Outline

1. Motivation
2. Set Theory
3. Leveraging Set Theory (Fun Part)

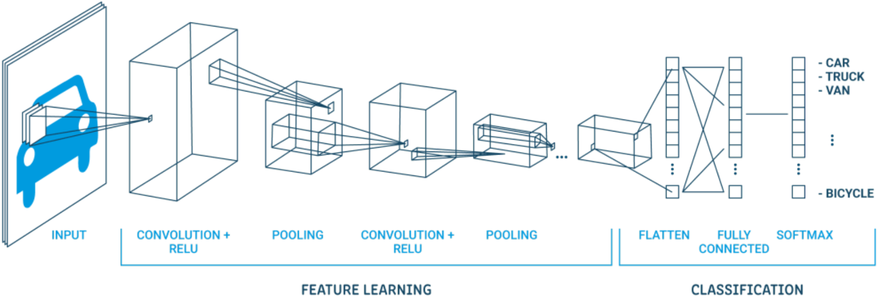
Introduction

Convolutional Neural Networks are *fantastic*. They efficiently extract a vast range of relevant contextual features and are resistant to pixel shift.



Introduction

Convolutional Neural Networks are *fantastic*. They efficiently extract a vast range of relevant contextual features and are resistant to pixel shift.



However, they have a **critical flaw**.

Neural Networks Aren't Rotationally Robust.

$\mathbf{Q}_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

Neural Networks Aren't Rotationally Robust.

$\mathbf{Q}_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

$\mathbf{A}_1$: Definitely!

Neural Networks Aren't Rotationally Robust.

Q$_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

A$_1$: Definitely!

Q$_2$: In practice, does this actually happen?

Neural Networks Aren't Rotationally Robust.

$\mathbf{Q}_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

$\mathbf{A}_1$: Definitely!

$\mathbf{Q}_2$: In practice, does this actually happen?
$\mathbf{A}_2$: Nope – all these images were misclassified.

Neural Networks Aren't Rotationally Robust.

Q$_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

A$_1$: Definitely!

Q$_2$: In practice, does this actually happen?
A$_2$: Nope – all these images were misclassified.

Q$_3$: How can we fix this?

Neural Networks Aren't Rotationally Robust.

Q$_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

A$_1$: Definitely!

Q$_2$: In practice, does this actually happen?
A$_2$: Nope – all these images were misclassified.

Q$_3$: How can we fix this?
A$_3$: Data Augmentation (boring)

Neural Networks Aren't Rotationally Robust.

$\mathbf{Q}_1$: Do you think that a CNN trained on a distribution of the left image *should* classify the right image as the same class for each of these pairs?

$\mathbf{A}_1$: Definitely!

$\mathbf{Q}_2$: In practice, does this actually happen?
$\mathbf{A}_2$: Nope – all these images were misclassified.

$\mathbf{Q}_3$: How can we fix this?
$\mathbf{A}_3$: Data Augmentation (boring), **G-Invariant Transformations** (fun)!

Outline

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$.

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \ \forall g_1 \in G$

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \ \forall g_1 \in G$
c. It contains an inverse. $\exists g_1^{-1} \in G$ s.t. $g_1 \odot g_1^{-1} = g_1^{-1} \odot g_1 = I$

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \ \forall g_1 \in G$
c. It contains an inverse. $\exists g_1^{-1} \in G$ s.t. $g_1 \odot g_1^{-1} = g_1^{-1} \odot g_1 = I$
d. $\odot$ is associative. $g_1 \odot (g_2 \odot g_3) = (g_1 \odot g_2) \odot g_3$

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \ \forall g_1 \in G$
c. It contains an inverse. $\exists g_1^{-1} \in G$ s.t. $g_1 \odot g_1^{-1} = g_1^{-1} \odot g_1 = I$
d. $\odot$ is associative. $g_1 \odot (g_2 \odot g_3) = (g_1 \odot g_2) \odot g_3$
e. (Optional) $\odot$ is commutative (abelian). $g_1 \odot g_2 = g_2 \odot g_1$

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \ \forall g_1 \in G$
c. It contains an inverse. $\exists g_1^{-1} \in G$ s.t. $g_1 \odot g_1^{-1} = g_1^{-1} \odot g_1 = I$
d. $\odot$ is associative. $g_1 \odot (g_2 \odot g_3) = (g_1 \odot g_2) \odot g_3$
e. (Optional) $\odot$ is commutative (abelian). $g_1 \odot g_2 = g_2 \odot g_1$
f. (Optional) Can be finite ($|G| < \infty$) or infinite ($|G| = \infty$).

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \; \forall g_1 \in G$
c. It contains an inverse. $\exists g_1^{-1} \in G$ s.t. $g_1 \odot g_1^{-1} = g_1^{-1} \odot g_1 = I$
d. $\odot$ is associative. $g_1 \odot (g_2 \odot g_3) = (g_1 \odot g_2) \odot g_3$
e. (Optional) $\odot$ is commutative (abelian). $g_1 \odot g_2 = g_2 \odot g_1$
f. (Optional) Can be finite ($|G| < \infty$) or infinite ($|G| = \infty$).

Q: Why do we care?

Understanding Groups

Definition: A group is a set G, with an operator $\odot$ that acts on $\forall g_1, g_2 \in G$. $(G, \odot)$ has the following properties:

a. It's closed under combination. $g_1 \odot g_2 = g_3 \in G$
b. It contains an identity. $\exists I \in G$ s.t. $g_1 \odot I = g_1 \; \forall g_1 \in G$
c. It contains an inverse. $\exists g_1^{-1} \in G$ s.t. $g_1 \odot g_1^{-1} = g_1^{-1} \odot g_1 = I$
d. $\odot$ is associative. $g_1 \odot (g_2 \odot g_3) = (g_1 \odot g_2) \odot g_3$
e. (Optional) $\odot$ is commutative (abelian). $g_1 \odot g_2 = g_2 \odot g_1$
f. (Optional) Can be finite ($|G| < \infty$) or infinite ($|G| = \infty$).

Q: Why do we care?
A: We leverage axioms a-d to derive a **transformation invariant representation** of our input. Invariance holds iff axioms a-d also hold.

The General Linear Group

A linear transformation is defined as:

$$T(v) = Av \text{ where } v \in \mathbb{R}^n,\ A \in \mathbb{R}^{m \times n},\ T : \mathbb{R}^n \to \mathbb{R}^m \tag{1}$$

The General Linear Group

A linear transformation is defined as:

$$T(v) = Av \text{ where } v \in \mathbb{R}^n,\ A \in \mathbb{R}^{m\times n},\ T : \mathbb{R}^n \to \mathbb{R}^m \tag{1}$$

If $A \in G$ is a linear transformation, G is a **linear group**.

The General Linear Group

A linear transformation is defined as:

$$T(v) = Av \text{ where } v \in \mathbb{R}^n,\ A \in \mathbb{R}^{m\times n},\ T : \mathbb{R}^n \to \mathbb{R}^m \tag{1}$$

If $A \in G$ is a linear transformation, G is a **linear group**.

The **general linear group** is the set of all invertible transformations:

$$GL_n : (M_{n\times n}(\mathbb{R}), \odot) \tag{2}$$

The General Linear Group

A linear transformation is defined as:

$$T(v) = Av \text{ where } v \in \mathbb{R}^n,\ A \in \mathbb{R}^{m \times n},\ T : \mathbb{R}^n \to \mathbb{R}^m \tag{1}$$

If $A \in G$ is a linear transformation, G is a **linear group**.

The **general linear group** is the set of all invertible transformations:

$$GL_n : (M_{n \times n}(\mathbb{R}), \odot) \tag{2}$$

Next, we define general linear groups over some affine transformations.

GL Tranformation Groups over Images

Let our input $x \in \mathbb{R}^{3\times n\times n}$ be our input image. Consider $\text{vec}(x) \in \mathbb{R}^{3n^2}$.

GL Tranformation Groups over Images

Let our input $x \in \mathbb{R}^{3 \times n \times n}$ be our input image. Consider $\text{vec}(x) \in \mathbb{R}^{3n^2}$.

$$G_{rot} \equiv \{T^{0^\circ}, T^{90^\circ}, T^{180^\circ}, T^{270^\circ}\} \tag{3}$$

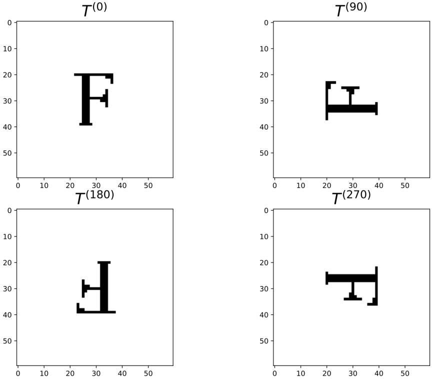
$$G_{flip} \equiv \{T^{v}, T^{h}, T^{180^\circ}, T^{0^\circ}\} \tag{4}$$

GL Tranformation Groups over Images

Let our input $x \in \mathbb{R}^{3 \times n \times n}$ be our input image. Consider $\text{vec}(x) \in \mathbb{R}^{3n^2}$.

$$G_{rot} \equiv \{T^{0^\circ}, T^{90^\circ}, T^{180^\circ}, T^{270^\circ}\} \quad (3)$$

$$G_{flip} \equiv \{T^{v}, T^{h}, T^{180^\circ}, T^{0^\circ}\} \quad (4)$$

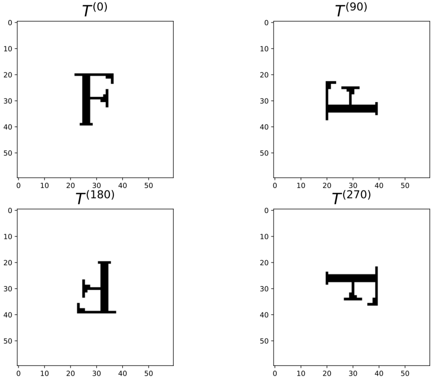


GL Tranformation Groups over Images

Let our input $x \in \mathbb{R}^{3 \times n \times n}$ be our input image. Consider $\text{vec}(x) \in \mathbb{R}^{3n^2}$.

$$G_{rot} \equiv \{T^{0^\circ}, T^{90^\circ}, T^{180^\circ}, T^{270^\circ}\} \tag{3}$$

$$G_{flip} \equiv \{T^{v}, T^{h}, T^{180^\circ}, T^{0^\circ}\} \tag{4}$$



Both groups are defined over $G : \mathbb{R}^{3n^2} \to \mathbb{R}^{3n^2}$

Outline

Obtaining an Invariant Transform

Defining the transformations as a **group** gives us *guarantees* we can exploit to ensure **invariance** to those transformations.

Obtaining an Invariant Transform

Defining the transformations as a **group** gives us *guarantees* we can exploit to ensure **invariance** to those transformations.

Formally, we want to make our input image invariant to rotation:

$$\forall T,\ \bar{T}(\boldsymbol{T}x) = \bar{T}x \text{ where } \boldsymbol{T} \in G_{rot} \tag{5}$$

Obtaining an Invariant Transform

Defining the transformations as a **group** gives us *guarantees* we can exploit to ensure **invariance** to those transformations.

Formally, we want to make our input image invariant to rotation:

$$\forall T,\ \bar{T}(\boldsymbol{T}x) = \bar{T}x \text{ where } \boldsymbol{T} \in G_{rot} \tag{5}$$

We can integrate this into the definition of a nueron:

$$\sigma(w^T x + b) \stackrel{def}{=} \sigma(w^T \boldsymbol{T} x + b) \tag{6}$$

Obtaining an Invariant Transform

Defining the transformations as a **group** gives us *guarantees* we can exploit to ensure **invariance** to those transformations.

Formally, we want to make our input image invariant to rotation:

$$\forall T,\ \bar{T}(\boldsymbol{T}x) = \bar{T}x \text{ where } \boldsymbol{T} \in G_{rot} \tag{5}$$

We can integrate this into the definition of a nueron:

$$\sigma(w^T x + b) \stackrel{def}{=} \sigma(w^T \boldsymbol{T} x + b) \tag{6}$$

Lemma: We can find $\bar{T}$ using the *Reynold's Operator*.

$$\bar{T} = \frac{1}{|G|} \sum_{g \in G} g \tag{7}$$

Obtaining an Invariant Transform

Defining the transformations as a **group** gives us *guarantees* we can exploit to ensure **invariance** to those transformations.

Formally, we want to make our input image invariant to rotation:

$$\forall T, \;\; \bar{T}(\boldsymbol{T}x) = \bar{T}x \text{ where } \boldsymbol{T} \in G_{rot} \tag{5}$$

We can integrate this into the definition of a nueron:

$$\sigma(w^T x + b) \stackrel{def}{=} \sigma(w^T \boldsymbol{T} x + b) \tag{6}$$

Lemma: We can find $\bar{T}$ using the *Reynold's Operator*.

$$\bar{T} = \frac{1}{|G|} \sum_{g \in G} g \tag{7}$$

Now, we have $\bar{T}$ s.t. $\bar{T} \circ T = T$!

Invariant Subspace of $\bar{T}$

Now, we need to find $M \subseteq \mathbb{R}^d$ s.t. $\forall w^T \in M,\ w^T \bar{T} \in M$.

Invariant Subspace of $\bar{T}$

Now, we need to find $M \subseteq \mathbb{R}^d$ s.t. $\forall w^T \in M,\ w^T \bar{T} \in M$.

One example of an invariant subspace is the left-1 eigenspace of $\bar{T}$:

$$\text{Left-1-Eig}(\bar{T}) = \{w \in \mathbb{R}^d \mid w_T \bar{T} = w^T\} \tag{8}$$

Invariant Subspace of $\bar{T}$

Now, we need to find $M \subseteq \mathbb{R}^d$ s.t. $\forall w^T \in M,\ w^T \bar{T} \in M$.

One example of an invariant subspace is the left-1 eigenspace of $\bar{T}$:

$$\text{Left-1-Eig}(\bar{T}) = \{w \in \mathbb{R}^d \mid w_T \bar{T} = w^T\} \tag{8}$$

Extracting those eigenvectors, we get $V = \{\boldsymbol{v}_i\}_{i=1}^k$ s.t. $\forall \boldsymbol{v}_i \in V$:

$$\boldsymbol{v}_i^T \bar{T} = \lambda_i \boldsymbol{v}_i \tag{9}$$

By definition of eigenvectors.

Invariant Subspace of $\bar{T}$

Now, we need to find $M \subseteq \mathbb{R}^d$ s.t. $\forall w^T \in M,\ w^T \bar{T} \in M$.

One example of an invariant subspace is the left-1 eigenspace of $\bar{T}$:

$$\text{Left-1-Eig}(\bar{T}) = \{w \in \mathbb{R}^d \mid w_T \bar{T} = w^T\} \tag{8}$$

Extracting those eigenvectors, we get $V = \{\boldsymbol{v}_i\}_{i=1}^k$ s.t. $\forall \boldsymbol{v}_i \in V$:

$$\boldsymbol{v}_i^T \bar{T} = \lambda_i \boldsymbol{v}_i \tag{9}$$

By definition of eigenvectors.

Since $\bar{T}$ is a projection operator, $\lambda_i = 1\ \forall i$:

$$\boldsymbol{v}_i^T \bar{T} = \boldsymbol{v}_i \tag{10}$$

Putting it All Together

We can use our invariant bases $V = \{\boldsymbol{v}_i\}_{i=1}^{k}$ to create an invariant layer.

$$w^T = \sum_{i=1}^{k} w_i \boldsymbol{v}_i \tag{11}$$

Putting it All Together

We can use our invariant bases $V = \{\boldsymbol{v}_i\}_{i=1}^{k}$ to create an invariant layer.

$$w^T = \sum_{i=1}^{k} w_i \boldsymbol{v}_i \tag{11}$$

Finally, we construct our group invariant layer:

$$h_{inv} = \sigma(w^T x + b) = \sigma(w^T T x + b) \; \forall T \in G_{rot} \tag{12}$$

Putting it All Together

We can use our invariant bases $V = \{\boldsymbol{v}_i\}_{i=1}^{k}$ to create an invariant layer.

$$w^T = \sum_{i=1}^{k} w_i \boldsymbol{v}_i \tag{11}$$

Finally, we construct our group invariant layer:

$$h_{inv} = \sigma(w^T x + b) = \sigma(w^T T x + b) \; \forall T \in G_{rot} \tag{12}$$

From here, the rest of the MLP follows the standard definition.

Equivariance for CNNs

Our previous approach leveraged an isomorphism, which requires $\mathbb{R}^n \to \mathbb{R}^n$. This fails to hold for CNNs.

Equivariance for CNNs

Our previous approach leveraged an isomorphism, which requires $\mathbb{R}^n \to \mathbb{R}^n$. This fails to hold for CNNs.

Instead, a more relevant property we can investigate is equivariance:

$$\rho_1(g)Wx = W\rho_2(g)x;\ g \in G, \rho_1 : G \to \mathbb{R}^{n\times n},\ \rho_2 : G \to \mathbb{R}^{k\times k} \tag{13}$$

Equivalence for CNNs

Our previous approach leveraged an isomorphism, which requires $\mathbb{R}^n \to \mathbb{R}^n$. This fails to hold for CNNs.

Instead, a more relevant property we can investigate is equivariance:

$$\rho_1(g)Wx = W\rho_2(g)x;\ g \in G, \rho_1 : G \to \mathbb{R}^{n\times n},\ \rho_2 : G \to \mathbb{R}^{k\times k} \tag{13}$$

Since this holds over our entire input x, we can re-arrange it as:

$$\rho_1(g)W\rho_2(g)^{-1} = W \tag{14}$$

Equivariance for CNNs

Our previous approach leveraged an isomorphism, which requires $\mathbb{R}^n \to \mathbb{R}^n$. This fails to hold for CNNs.

Instead, a more relevant property we can investigate is equivariance:

$$\rho_1(g)Wx = W\rho_2(g)x; \; g \in G, \rho_1 : G \to \mathbb{R}^{n \times n}, \; \rho_2 : G \to \mathbb{R}^{k \times k} \tag{13}$$

Since this holds over our entire input x, we can re-arrange it as:

$$\rho_1(g)W\rho_2(g)^{-1} = W \tag{14}$$

This can be re-arranged to demonstrate an equivalency with *invariance*:

$$\underbrace{\rho_2(g) \times \rho_1(g^{-1})^T}_{\bar{T}} \underbrace{\text{vec}(W)}_{x} = \underbrace{\text{vec}(W)}_{x} \tag{15}$$

Equivariance for CNNs

Our previous approach leveraged an isomorphism, which requires $\mathbb{R}^n \to \mathbb{R}^n$. This fails to hold for CNNs.

Instead, a more relevant property we can investigate is equivariance:

$$\rho_1(g)Wx = W\rho_2(g)x;\ g \in G, \rho_1 : G \to \mathbb{R}^{n\times n},\ \rho_2 : G \to \mathbb{R}^{k\times k} \tag{13}$$

Since this holds over our entire input x, we can re-arrange it as:

$$\rho_1(g)W\rho_2(g)^{-1} = W \tag{14}$$

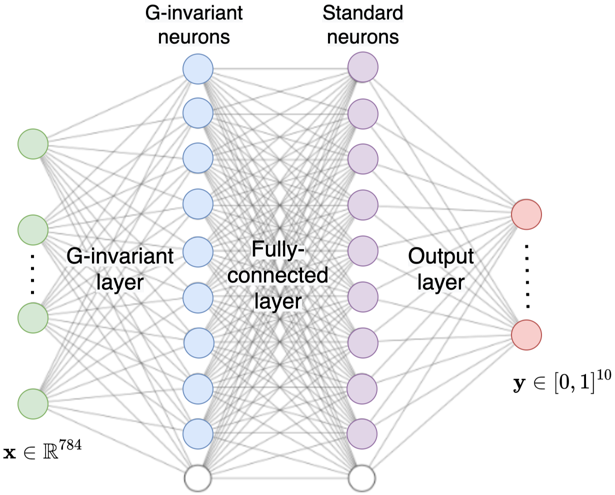
This can be re-arranged to demonstrate an equivalency with *invariance*:

$$\underbrace{\rho_2(g) \times \rho_1(g^{-1})^T}_{\bar{T}}\underbrace{\text{vec}(W)}_{x} = \underbrace{\text{vec}(W)}_{x} \tag{15}$$

From there, the previous invariance proof follows.

Let's Demonstrate!

Here's what the final architecture looks like:



Thank you!

Have an awesome rest of your day!

Paper: `https://arxiv.org/abs/2104.10105/`

Slides: `https://cs.purdue.edu/homes/jsetpal/slides/gti.pdf`

Notebook: `https://cs.purdue.edu/homes/jsetpal/nb/gti.ipynb`