# Grassmannian Dimensionality Reduction Using Triplet Margin Loss for Universal Manifold Embedding Classification of 3D Point Clouds

Yuval Haitman*, Joseph M. Francos* and Louis L. Scharf†

*Electrical & Computer Engineering Department, Ben-Gurion University, Beer-Sheva, Israel
† Department of Mathematics, Colorado State University, Fort Collins, CO, USA

April 18, 2022

# Detection and Classification of 3-D Objects Undergoing Rigid Transformations

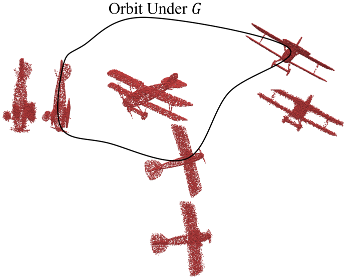
- Consider a 3-D object $s \in \{s_1, \cdots, s_K\}$, and the *orbit* of equivalent observations formed by the action of the transformation group $G = SE(3)$ on $s$.

# Detection and Classification of 3-D Objects Undergoing Rigid Transformations

- Consider a 3-D object $s \in \{s_1, \cdots, s_K\}$, and the *orbit* of equivalent observations formed by the action of the transformation group $G = SE(3)$ on $s$.
- The set of possible observations on these equivalent objects is generally a manifold in the ambient space of observations.

# Detection and Classification of 3-D Objects Undergoing Rigid Transformations

- Consider a 3-D object $s \in \{s_1, \cdots, s_K\}$, and the *orbit* of equivalent observations formed by the action of the transformation group $G = SE(3)$ on $s$.
- The set of possible observations on these equivalent objects is generally a manifold in the ambient space of observations.
- In the presence of observation noise and random sampling patterns of the point clouds, the observations do not lie strictly on the manifold.



# RTUME for Classification

- The Rigid Transformation Universal Manifold Embedding (RTUME)[1] provides a mapping from the orbit of observations on some object to a single low dimensional linear subspace of Euclidean space.

[1] Amit Efraim and Joseph M Francos. "The Universal Manifold Embedding for Estimating Rigid Transformations of Point Clouds". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 5157–5161.

# RTUME for Classification

- The Rigid Transformation Universal Manifold Embedding (RTUME)[1] provides a mapping from the orbit of observations on some object to a single low dimensional linear subspace of Euclidean space.
- This linear subspace is invariant to the geometric transformations and hence is a representative of the orbit.

[1] Amit Efraim and Joseph M Francos. "The Universal Manifold Embedding for Estimating Rigid Transformations of Point Clouds". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 5157–5161.

# RTUME for Classification

- The Rigid Transformation Universal Manifold Embedding (RTUME)[1] provides a mapping from the orbit of observations on some object to a single low dimensional linear subspace of Euclidean space.
- This linear subspace is invariant to the geometric transformations and hence is a representative of the orbit.
- In the classification set-up the RTUME subspace extracted from an experimental observation is tested against a set of subspaces representing the different object manifolds, in search for the nearest class.

[1] Amit Efraim and Joseph M Francos. "The Universal Manifold Embedding for Estimating Rigid Transformations of Point Clouds". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 5157–5161.

# Rigid Transformation Universal Manifold Embedding (RTUME)

- Let $h(\mathbf{x}), g(\mathbf{x})$ be two observations on the same object related by a rigid transformation:
$$h(\mathbf{x}) = g(\mathbf{R}\mathbf{x} + \mathbf{t}) \tag{1}$$
where $h(\mathbf{x}), g(\mathbf{x})$ are evaluated from the raw point cloud measurements using an SE$(3)$-invariant function.
- We use the matrix representation of SE$(3)$ in homogeneous coordinates with right multiplication:
$$\mathbf{D}(\mathbf{R}, \mathbf{t}) = \begin{bmatrix} 1 & \mathbf{t}^T \\ \mathbf{0} & \mathbf{R}^T \end{bmatrix} \tag{2}$$

# RTUME - Matrix Representation

## RTUME Matrix

$$\mathbf{T}(h) = \begin{bmatrix} \int\limits_{\mathbb{R}^3} w_1 \circ h(\mathbf{x}) d\mathbf{x} & \int\limits_{\mathbb{R}^3} x_1 w_1 \circ h(\mathbf{x}) d\mathbf{x} & \dots & \int\limits_{\mathbb{R}^3} x_3 w_1 \circ h(\mathbf{x}) d\mathbf{x} \\ & \vdots & & \\ \int\limits_{\mathbb{R}^3} w_M \circ h(\mathbf{x}) d\mathbf{x} & \int\limits_{\mathbb{R}^3} x_1 w_M \circ h(\mathbf{x}) d\mathbf{x} & \dots & \int\limits_{\mathbb{R}^3} x_3 w_M \circ h(\mathbf{x}) d\mathbf{x} \end{bmatrix} \tag{3}$$

- $\{w_m\}_{m=1}^{M}$ are measurable functions aimed at generating many compandings of the observation.

# RTUME - Matrix Representation

## RTUME Matrix

$$\mathbf{T}(h) = \begin{bmatrix} \int\limits_{\mathbb{R}^3} w_1 \circ h(\mathbf{x}) d\mathbf{x} & \int\limits_{\mathbb{R}^3} x_1 w_1 \circ h(\mathbf{x}) d\mathbf{x} & \dots & \int\limits_{\mathbb{R}^3} x_3 w_1 \circ h(\mathbf{x}) d\mathbf{x} \\ & \vdots & & \\ \int\limits_{\mathbb{R}^3} w_M \circ h(\mathbf{x}) d\mathbf{x} & \int\limits_{\mathbb{R}^3} x_1 w_M \circ h(\mathbf{x}) d\mathbf{x} & \dots & \int\limits_{\mathbb{R}^3} x_3 w_M \circ h(\mathbf{x}) d\mathbf{x} \end{bmatrix} \quad (3)$$

- $\{w_m\}_{m=1}^{M}$ are measurable functions aimed at generating many compandings of the observation.
- The RTUME matrices of $h(\mathbf{x}), g(\mathbf{x})$ are related by:

$$\mathbf{T}(h) = \mathbf{T}(g)\mathbf{D}^{-1}(\mathbf{R}, \mathbf{t}) \quad (4)$$

- Since $\mathbf{T}(h)$ and $\mathbf{T}(g)$ are related by a right invertible linear transformation, the column space of $\mathbf{T}(g)$ and the column space of $\mathbf{T}(h)$ are identical.

# Design of the RTUME Operator: TL-GDRUME

- Classifier performance highly depends on the choice of the set of functions composing the UME operator.
- Find the functions, that best separates the RTUME representation of each object from those of the other objects, while minimizing the distance between observations on the same object.

$$d_{pF}(\langle \mathbf{T}(Z) \rangle, \langle \mathbf{T}(X) \rangle) = \frac{1}{\sqrt{2}} ||\mathbf{P}_X - \mathbf{P}_Z||_F = ||\sin \boldsymbol{\theta}||_2 \tag{5}$$

# Design of the RTUME Operator: TL-GDRUME

- Classifier performance highly depends on the choice of the set of functions composing the UME operator.
- Find the functions, that best separates the RTUME representation of each object from those of the other objects, while minimizing the distance between observations on the same object.

$$d_{pF}(\langle \mathbf{T}(Z) \rangle, \langle \mathbf{T}(X) \rangle) = \frac{1}{\sqrt{2}} ||\mathbf{P}_X - \mathbf{P}_Z||_F = ||\sin \boldsymbol{\theta}||_2 \tag{5}$$

- Using Grassmannian dimensionality reduction and metric learning scheme we derive TL-GDRUME: An analytic solution for designing the RTUME operators.

# Level Set Representation

- Given an observation $X(\mathbf{u}), \mathbf{u} \in \mathbb{R}^3$ the level-set representation of $X(\mathbf{u})$is:

$$X(\mathbf{u}) = \sum_{i=1}^{Q} q_i \mathcal{I}_i^X(\mathbf{u}) \tag{6}$$

where $\mathcal{I}_i^X(\mathbf{u})$ is the indicator function of the level-set of $\mathbf{u}$ where $q_{i-1} \leq X(\mathbf{u}) \leq q_i$.

# Level Set Representation

- Given an observation $X(\mathbf{u}), \mathbf{u} \in \mathbb{R}^3$ the level-set representation of $X(\mathbf{u})$is:

$$X(\mathbf{u}) = \sum_{i=1}^{Q} q_i \mathcal{I}_i^X(\mathbf{u}) \tag{6}$$

where $\mathcal{I}_i^X(\mathbf{u})$ is the indicator function of the level-set of $\mathbf{u}$ where $q_{i-1} \leq X(\mathbf{u}) \leq q_i$.

- The action of a measurable function $w_m$ on the level-set representation of $X(\mathbf{u})$ is to map $q_i$ to $w_m(q_i)$:

$$w_m(X(\mathbf{u})) = \sum_{i=1}^{Q} w_m(q_i) \mathcal{I}_i^X(\mathbf{u}) \tag{7}$$

# Fundamental Universal Manifold Embedding (FUME)

- Using the level-set representation of $X(\mathbf{u})$ each term of the RTUME matrix $\mathbf{T}(X)$ becomes:

$$\mathbf{T}_{m,j} = \int_{\mathbb{R}^3} w_m \circ X(\mathbf{u}) u_j d\mathbf{u} = \sum_{i=1}^{Q} w_m(q_i) \underbrace{\int_{\mathbb{R}^3} \mathcal{I}_i^X(\mathbf{u}) u_j d\mathbf{u}}_{\mathbf{F}_{i,j}^X} = \sum_{i=1}^{Q} w_{m,i} \mathbf{F}_{i,j}^X \quad (8)$$

# Fundamental Universal Manifold Embedding (FUME)

- Using the level-set representation of $X(\mathbf{u})$ each term of the RTUME matrix $\mathbf{T}(X)$ becomes:

$$\mathbf{T}_{m,j} = \int_{\mathbb{R}^3} w_m \circ X(\mathbf{u}) u_j d\mathbf{u} = \sum_{i=1}^{Q} w_m(q_i) \underbrace{\int_{\mathbb{R}^3} \mathcal{I}_i^X(\mathbf{u}) u_j d\mathbf{u}}_{\mathbf{F}_{i,j}^X} = \sum_{i=1}^{Q} w_{m,i} \mathbf{F}_{i,j}^X \quad (8)$$

- and

$$\mathbf{T}(X) = \mathbf{W}^T \mathbf{F}^X; \qquad \mathbf{W}^T = \{w_{m,i}\} \in \mathbb{R}^{M \times Q} \quad (9)$$

# Fundamental Universal Manifold Embedding (FUME)

- Using the level-set representation of $X(\mathbf{u})$ each term of the RTUME matrix $\mathbf{T}(X)$ becomes:

$$\mathbf{T}_{m,j} = \int_{\mathbb{R}^3} w_m \circ X(\mathbf{u}) u_j d\mathbf{u} = \sum_{i=1}^{Q} w_m(q_i) \underbrace{\int_{\mathbb{R}^3} \mathcal{I}_i^X(\mathbf{u}) u_j d\mathbf{u}}_{\mathbf{F}_{i,j}^X} = \sum_{i=1}^{Q} w_{m,i} \mathbf{F}_{i,j}^X \quad (8)$$

- and

$$\mathbf{T}(X) = \mathbf{W}^T \mathbf{F}^X; \qquad \mathbf{W}^T = \{w_{m,i}\} \in \mathbb{R}^{M \times Q} \quad (9)$$

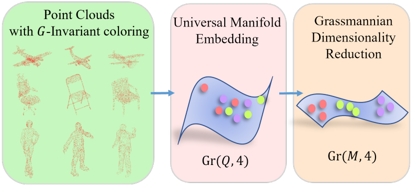
- $\mathbf{F}^X = \{\mathbf{F}_{i,j}^X\} \in \mathbb{R}^{M \times Q}$ is the *Fundamental Universal Manifold Embedding* (FUME) matrix of $X(\mathbf{u})$.

- Since $M \leq Q$ the role of $\mathbf{W}$ is to transform the subspace $\langle \mathbf{F}^X \rangle \in \mathrm{Gr}(Q, 4)$to the subspace $\langle \mathbf{G}^X \rangle \in \mathrm{Gr}(M, 4)$.

# Grassmannian Dimensionality Reduction

- Find $\mathbf{W} \in \mathbb{R}^{Q \times M}$ that jointly maps FUME subspaces from a Grassmannian with higher ambient space dimension to a Grassmannian with lower ambient space dimension.

$$\{\langle \mathbf{F}^k \rangle\}_{k=1}^N \in \mathsf{Gr}(Q,4) \xrightarrow{\mathbf{Y}^k = \mathbf{W}^T \mathbf{F}^k} \{\langle \mathbf{Y}^k \rangle\}_{k=1}^N \in \mathsf{Gr}(M,4) \quad (10)$$

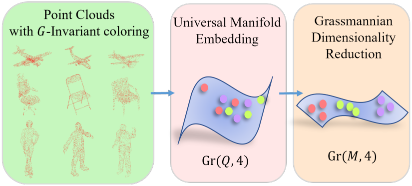


# Grassmannian Dimensionality Reduction

- Find $\mathbf{W} \in \mathbb{R}^{Q \times M}$ that jointly maps FUME subspaces from a Grassmannian with higher ambient space dimension to a Grassmannian with lower ambient space dimension.

$$\{\langle \mathbf{F}^k \rangle\}_{k=1}^{N} \in \mathsf{Gr}(Q,4) \xrightarrow{\mathbf{Y}^k = \mathbf{W}^T \mathbf{F}^k} \{\langle \mathbf{Y}^k \rangle\}_{k=1}^{N} \in \mathsf{Gr}(M,4) \quad (10)$$

- $\mathbf{W}$ is designed such that observations from the same orbit generate close together subspaces while those from different orbits generate far apart subspaces.

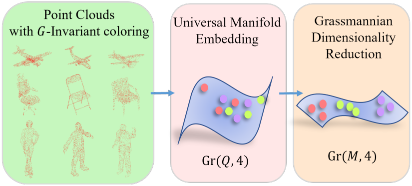


# Grassmannian Dimensionality Reduction

- Find $\mathbf{W} \in \mathbb{R}^{Q\times M}$ that jointly maps FUME subspaces from a Grassmannian with higher ambient space dimension to a Grassmannian with lower ambient space dimension.

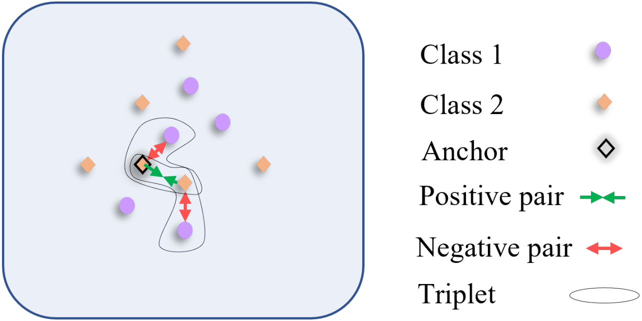
$$\{\langle \mathbf{F}^k \rangle\}_{k=1}^{N} \in \mathsf{Gr}(Q,4) \xrightarrow{\mathbf{Y}^k = \mathbf{W}^T\mathbf{F}^k} \{\langle \mathbf{Y}^k \rangle\}_{k=1}^{N} \in \mathsf{Gr}(M,4) \quad (10)$$

- $\mathbf{W}$ is designed such that observations from the same orbit generate close together subspaces while those from different orbits generate far apart subspaces.
- A sufficient condition that guarantees all $\langle \mathbf{Y}^k \rangle$ are indeed on $\mathsf{Gr}(M,4)$ is that $\mathbf{W}$ has a full column rank, or alternatively, that the columns of $\mathbf{W}$ are orthonormal.



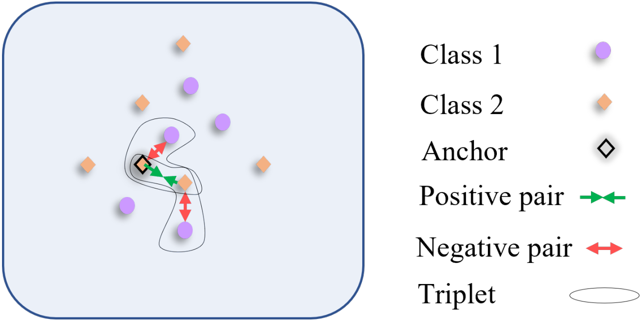
# Metric Learning - Triplet Margin Loss

- We employ metric learning with hard-negative mining.



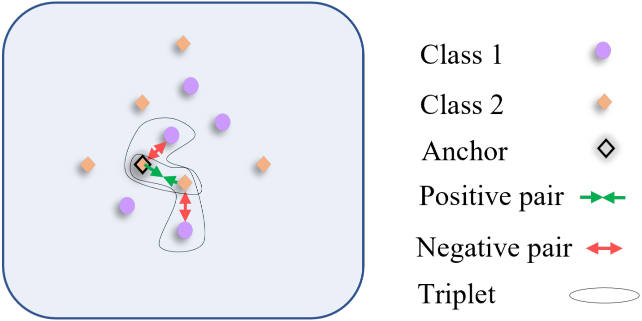
# Metric Learning - Triplet Margin Loss

- We employ metric learning with hard-negative mining.
- Triplet loss jointly minimizes the distance between a given anchor and its positive match, while maximizing the distance to the hardest negative example.



# Metric Learning - Triplet Margin Loss

- We employ metric learning with hard-negative mining.
- Triplet loss jointly minimizes the distance between a given anchor and its positive match, while maximizing the distance to the hardest negative example.
- Negative mining is applied both to the anchor and to its positive match.



# TL-GDRUME Optimization of $\mathbf{W}$

- Given a training set of $N$ labeled observations $\{X_i, k_i\}_{i=1}^N$, $k_i \in \{1, \ldots K\}$ and their corresponding FUME matrices $\{\mathbf{F}_i\}_{i=1}^N$.

# TL-GDRUME Optimization of $\mathbf{W}$

- Given a training set of $N$ labeled observations $\{X_i, k_i\}_{i=1}^{N}$, $k_i \in \{1, \ldots K\}$ and their corresponding FUME matrices $\{\mathbf{F}_i\}_{i=1}^{N}$.
- Examples are paired to positive ($\mathcal{P}$) and negative ($\mathcal{N}$) sets according to their class label, each set contains the indices of the appropriate examples.

# TL-GDRUME Optimization of $\mathbf{W}$

- Given a training set of $N$ labeled observations $\{X_i, k_i\}_{i=1}^N$, $k_i \in \{1, \dots K\}$ and their corresponding FUME matrices $\{\mathbf{F}_i\}_{i=1}^N$.
- Examples are paired to positive ($\mathcal{P}$) and negative ($\mathcal{N}$) sets according to their class label, each set contains the indices of the appropriate examples.
- Points on the Grassmann manifold are represented by an orthogonal basis of the RTUME matrices $\{\mathbf{W}^T\mathbf{F}_i\}_{i=1}^N$ using *QR- decomposition*.

# TL-GDRUME Optimization of $\mathbf{W}$

- Given a training set of $N$ labeled observations $\{X_i, k_i\}_{i=1}^{N}$, $k_i \in \{1, \ldots K\}$ and their corresponding FUME matrices $\{\mathbf{F}_i\}_{i=1}^{N}$.
- Examples are paired to positive ($\mathcal{P}$) and negative ($\mathcal{N}$) sets according to their class label, each set contains the indices of the appropriate examples.
- Points on the Grassmann manifold are represented by an orthogonal basis of the RTUME matrices $\{\mathbf{W}^T\mathbf{F}_i\}_{i=1}^{N}$ using *QR- decomposition*.
- Distance between examples is measured by the projection Frobenius-norm on the Grassmannian:

$$D_{i,j}(\mathbf{W}) = d_{pF}^2\left(\langle\mathbf{Q}_i(\mathbf{W})\rangle, \langle\mathbf{Q}_j(\mathbf{W})\rangle\right) \tag{11}$$

# TL-GDRUME Optimization of $\mathbf{W}$

## TL-GDRUME Training

$$\min_{\mathbf{W}\in\mathbb{R}^{Q\times M}} L(\mathbf{W}) = \sum_{(i,j)\in\mathcal{P}} \left[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{i,k}(\mathbf{W})\right]_+ + \left[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{j,k}(\mathbf{W})\right]_+ \tag{12}$$

subject to $\mathbf{W}^{\mathbf{T}}\mathbf{W} = \mathbf{I}_M$

- Since $\langle \mathbf{W}^T\mathbf{F}_i\rangle \in \mathrm{Gr}(M,4)$, distance values are bounded $D_{i,j}(\mathbf{W}) \in [0,4]$ therefore a typical value for the margin $m$ will be in this range.

[2] Nicolas Boumal et al. "Manopt, a Matlab toolbox for optimization on manifolds". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 1455–1459.

# TL-GDRUME Optimization of $\mathbf{W}$

## TL-GDRUME Training

$$\min_{\mathbf{W}\in\mathbb{R}^{Q\times M}} L(\mathbf{W}) = \sum_{(i,j)\in\mathcal{P}} \left[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{i,k}(\mathbf{W})\right]_+ + \left[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{j,k}(\mathbf{W})\right]_+ \tag{12}$$

subject to $\mathbf{W}^{\mathbf{T}}\mathbf{W} = \mathbf{I}_M$

- Since $\langle \mathbf{W}^T\mathbf{F}_i\rangle \in \mathrm{Gr}(M,4)$, distance values are bounded $D_{i,j}(\mathbf{W}) \in [0,4]$ therefore a typical value for the margin $m$ will be in this range.
- $[\cdot]_+ = \max(0,\cdot)$.

[2] Nicolas Boumal et al. "Manopt, a Matlab toolbox for optimization on manifolds". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 1455–1459.

# TL-GDRUME Optimization of $\mathbf{W}$

## TL-GDRUME Training

$$\min_{\mathbf{W}\in\mathbb{R}^{Q\times M}} L(\mathbf{W}) = \sum_{(i,j)\in\mathcal{P}} \left[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{i,k}(\mathbf{W})\right]_+ + \left[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{j,k}(\mathbf{W})\right]_+ \tag{12}$$

subject to $\mathbf{W}^{\mathbf{T}}\mathbf{W} = \mathbf{I}_M$

- Since $\langle \mathbf{W}^T\mathbf{F}_i\rangle \in \mathrm{Gr}(M,4)$, distance values are bounded $D_{i,j}(\mathbf{W}) \in [0,4]$ therefore a typical value for the margin $m$ will be in this range.
- $[\cdot]_+ = \max(0,\cdot)$.
- We solve an optimization problem on the Stiefel manifold (12) using manifold optimization toolbox Manopt[2].

[2] Nicolas Boumal et al. "Manopt, a Matlab toolbox for optimization on manifolds". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 1455–1459.

# Experimental Setup and Results

- Evaluation on ModelNet40 point cloud dataset.

| Sampling | Method | 0.5 MR noise | 0.8 MR noise |
|---|---|---|---|
| Uniform | FUME | 0.85 | 0.83 |
| | TL-GDRUME | **0.93** | **0.91** |
| Non - Uniform | FUME | 0.83 | 0.81 |
| | TL-GDRUME | **0.92** | **0.90** |

Table: Accuracy comparison of FUME and TL-GDRUME on deformed ModelNet40 observations, uniformly and non-uniformly sampled, in the presence of noise.

# Experimental Setup and Results

- Evaluation on ModelNet40 point cloud dataset.
- For each class we generated observations that differ by rigid transformation, additive noise and random sampling.

| Sampling | Method | 0.5 MR noise | 0.8 MR noise |
|---|---|---|---|
| Uniform | FUME | 0.85 | 0.83 |
| | TL-GDRUME | **0.93** | **0.91** |
| Non - Uniform | FUME | 0.83 | 0.81 |
| | TL-GDRUME | **0.92** | **0.90** |

Table: Accuracy comparison of FUME and TL-GDRUME on deformed ModelNet40 observations, uniformly and non-uniformly sampled, in the presence of noise.

# Experimental Setup and Results

- Evaluation on ModelNet40 point cloud dataset.
- For each class we generated observations that differ by rigid transformation, additive noise and random sampling.
- The class of each observation in the test is determined to be the label of its nearest neighbor in the training set.

| Sampling | Method | 0.5 MR noise | 0.8 MR noise |
|---|---|---|---|
| Uniform | FUME | 0.85 | 0.83 |
| | TL-GDRUME | **0.93** | **0.91** |
| Non - Uniform | FUME | 0.83 | 0.81 |
| | TL-GDRUME | **0.92** | **0.90** |

Table: Accuracy comparison of FUME and TL-GDRUME on deformed ModelNet40 observations, uniformly and non-uniformly sampled, in the presence of noise.

# Experimental Setup and Results

- Evaluation on ModelNet40 point cloud dataset.
- For each class we generated observations that differ by rigid transformation, additive noise and random sampling.
- The class of each observation in the test is determined to be the label of its nearest neighbor in the training set.
- We tested the classification performance under two different noise statistics and two sampling methods - uniform and non-uniform.

| Sampling | Method | 0.5 MR noise | 0.8 MR noise |
|---|---|---|---|
| Uniform | FUME | 0.85 | 0.83 |
| | TL-GDRUME | **0.93** | **0.91** |
| Non - Uniform | FUME | 0.83 | 0.81 |
| | TL-GDRUME | **0.92** | **0.90** |

Table: Accuracy comparison of FUME and TL-GDRUME on deformed ModelNet40 observations, uniformly and non-uniformly sampled, in the presence of noise.

# Conclusions

- We have presented a novel approach for designing the RTUME of 3D point clouds towards optimizing its performance for detection and classification tasks.
- In the presence of observation noise and challenging sampling patterns, the observations do not lie strictly on the manifold and the resulting RTUME subspaces are noisy. Yet, TL-GDRUME provides highly accurate classification results compared to the naive FUME.

*Thank You*