Swish-Driven GoogleNet for Intelligent Analog Beam Selection in Terahertz Beamspace MIMO

Hosein Zarini†, Mohammad Robat Mili§, Mehdi Rasti†,**, Sergey Andreev*, and Pedro H. J. Nardelli**
†Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran
§Electronics Research Institute, Sharif University of Technology, Tehran, Iran
*Tampere University, Tampere, Finland
**Lappeenranta-Lahti University of Technology, Lappeenranta, Finland

Abstract—**In this paper, we propose an intelligent analog beam selection strategy in a terahertz (THz) band beamspace multiple-input multiple-output (MIMO) system. First inspired by transfer learning, we fine-tune the pre-trained off-the-shelf GoogleNet classifier to learn analog beam selection as a multi-class mapping problem. Simulation results show 83% accuracy for the analog beam selection, which subsequently results in 12% spectral efficiency (SE) gain over the existing counterparts. For a more accurate classifier, we replace the conventional rectified linear unit (ReLU) activation function of the GoogleNet with the recently proposed Swish and retrain the fine-tuned GoogleNet to learn analog beam selection. It is numerically indicated that the fine-tuned Swish-driven GoogleNet achieves 86% accuracy, as well as 18% improvement in achievable SE, over the similar schemes. Eventually, a strong ensembled classifier is developed to learn analog beam selection by sequentially training multiple fine-tuned Swish-driven GoogleNet classifiers. According to the simulations, the strong ensembled model is 90% accurate and yields 27% gain in achievable SE in comparison with prior methods.**

Index Terms—**Terahertz (THz) band, beamspace, multiple-input multiple-output, analog beam selection, GoogleNet, Swish, ensembled classifier.**

I. Introduction

Over the recent years, beamspace technology [1] has attracted much attention in high-frequency bands, as an alternative to the conventional massive multiple-input-multiple-output (MIMO) architecture. In the latter case, each antenna element requires a specific radio frequency (RF) chain[1], which makes this architecture inefficient in practice, owing to a massive number of required RF chains. In beamspace technology though, the scattered signals of divergent paths (beams) can be concentrated on a limited number of dominant beams and the spatial domain channel is thereby transformed into the beamspace domain channel. For this reason, from a massive number of beams, merely a limited number is adopted, which in turn necessitates fewer RF chains for reliable beam steering.

Hybrid analog-digital beamspace MIMO is consequently a reasonable system in terms of energy, cost, and complexity, provided that analog beam selection is efficiently performed. However, this sets out new challenges due to the massive number of beams. While on the one hand, the prior optimization-based analog beam selection efforts such as those in [2] impose expensive computational burden on the transceivers, the low-complexity machine/deep learning approaches like [3] and [4] on the other hand, suffer from accuracy loss in this regard. According to the statistics in [5], trained on environmental samples (e.g., the line-of-sight (LoS) and non-line-of-sight (NLoS) beams), two well-known classifiers i.e., the linear SVM [3] and the decision tree [4] are only 33% and 55% accurate, respectively, which in turn brings about a non-negligible performance loss for the beamspace architecture.

The main contribution of this paper is to mitigate the precision drop in prior learning-aided works on analog beam selection by proposing a fine-tuned deep learning technique, along with an ensemble learning technique as follows.

- First, we consider the analog beam selection problem as a multi-class classification task. To this aim, we retrain the pre-trained off-the-shelf GoogleNet classifier [6] based on the concept of transfer learning [7], so as to learn the analog beam selection. Simulation results verify that the retrained GoogleNet exhibits some 83% accuracy for the analog beam selection and offers up to 12% gain in achievable spectral efficiency (SE) over the counterparts, if the signal-to-noise-ratio (SNR) is 30dB.
- We fine-tune the GoogleNet classifier for further precision by replacing its conventional activation function i.e., the rectified linear unit (ReLU) with the Swish activation function [8]. It is numerically shown that retraining the fine-tuned GoogleNet achieves some 86% accuracy, as well as 18% achievable SE gain over the counterparts, at SNR = 30dB.
- In addition, the performance of the proposed analog beam selection scheme is further enhanced by sequentially incorporating a set of fine-tuned GoogleNets (each one is known as a weak learner) into an ensembled model (known as a strong learner) [9]. The designed strong learner according to the simulations exceeds the achievable SE of the prior counterparts by up to 27%, while yielding 90% accuracy, at SNR = 30dB.

In remainder of the paper, Sections II and III describe the system setup and the solution approach, whereas the simulation results and conclusions are presented in Sections IV and V, respectively.

[1] RF chains are known as dominant modules in energy consumption, hardware cost, and complexity order of conventional massive MIMO systems.

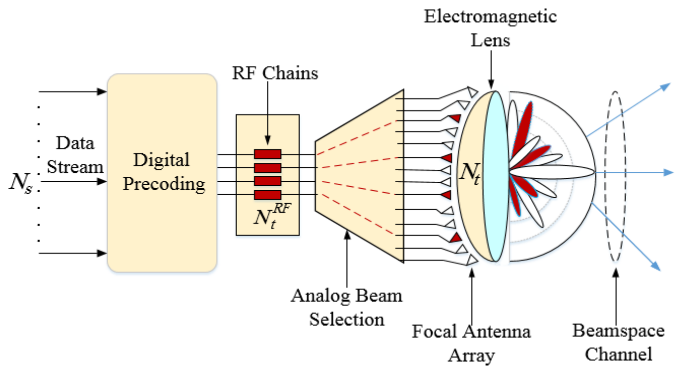


Fig. 1: Hybrid analog-digital beamspace MIMO architecture at the transmitter.

II. System Setup

A. Hybrid Analog-Digital Architecture

Consider the downlink THz communication, where the transmitter employs $N_{\rm t}$ ($N_{\rm t}^{RF}$) transmit antennas (transmit RF chains) for serving a receiver equipped with $N_{\rm r}$ ($N_{\rm r}^{RF}$) receive antennas (receive RF chains). The power-normalized transmit symbols are denoted by $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$, where $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{I}_{N_s}$. The transceivers employ a hybrid analog-digital beamspace architecture to preserve the system flexibility, as well as the efficiency in hardware cost, and energy consumption [1]. As demonstrated, in Fig. 1, a baseband digital matrix $\mathbf{F}_{\rm BB} \in \mathbb{C}^{N_{\rm t}^{RF} \times N_s}$ is leveraged at the transmitter, followed by an analog beam selection network denoted by $\mathbf{S}_{\rm t} \in \mathbb{R}^{N_{\rm t} \times N_{\rm t}^{RF}}$ in matrix form for mapping $N_{\rm t}^{RF}$ transmit RF chains onto a subset of $N_{\rm t}$ transmit antennas/beams. Further, a lens antenna array is deployed at the transmitter, including an energy-focusing electromagnetic lens, where its focal surface is equipped with a large-scale antenna array.

At the receiver side, once the lens antenna array receives the signals, a mapping is performed between the predominant receive antennas/beams and the receive RF chains through the receive analog beam selection network $\mathbf{S}_{\rm r} \in \mathbb{R}^{N_{\rm r} \times N_{\rm r}^{RF}}$, where a baseband digital combining matrix $\mathbf{W}_{\rm BB} \in \mathbb{C}^{N_{\rm r}^{RF} \times N_s}$ is embedded afterwards to obtain the transmit symbols. Hence, the discrete-time received baseband complex signal is given by $y = \mathbf{W}_{\rm BB}^H \mathbf{S}_{\rm r}^H \mathbf{H}_b \mathbf{x} + \mathbf{W}_{\rm BB}^H \mathbf{S}_{\rm r}^H \mathbf{n}$, wherein $\mathbf{n} \sim N(0, \sigma^2 \mathbf{I}_{N_{\rm r}})$ is the additive white Gaussian noise (AWGN) with the noise power σ^2 and $\mathbf{H}_b$ denotes the THz beamspace channel.

B. Spatial Domain THz Channel

According to the well-known Saleh-Valenzuela geometric model [11], a ray-based clustered THz channel is assumed to have N_{cl} cluster of scatterers, each contributing N_{ray} propagation rays. Further, a limited angle-of-departure/arrival (AoD/AoA) spread is supposed for a typical cluster l, denoted by $\psi_{\rm t}^l$ and $\psi_{\rm r}^l$, respectively. For a typical cluster/ray l/u, the complex-valued gain is denoted by $\alpha^{l,u}$, while the physical AoD and AoA for the transmitter and the receiver are respectively denoted by $\theta_{\rm t}^{l,u} \in \psi_{\rm t}^l$, and $\theta_{\rm r}^{l,u} \in \psi_{\rm r}^l$. Let us denote the antenna element spacing by d, the speed of light by c, the wavelength by $\lambda = c/f_c$, and the carrier frequency by f_c. Then, the spatial AoD/AoA can be represented by $\phi_{\rm t}^{l,u} = (d/\lambda)\sin\theta_{\rm t}^{l,u}$ and $\phi_{\rm r}^{l,u} = (d/\lambda)\sin\theta_{\rm r}^{l,u}$, respectively. Accordingly, the narrow-band discrete-time spatial domain THz channel $\mathbf{H} \in \mathbb{C}^{N_{\rm r} \times N_{\rm t}}$ is expressed as $\mathbf{H} = \gamma \sum_{l=1}^{N_{cl}} \sum_{u=1}^{N_{ray}} \alpha_{l,u} \mathbf{a}_{\rm r}\left(\phi_{\rm r}^{l,u}\right) \mathbf{a}_{\rm t}^{\rm H}\left(\phi_{\rm t}^{l,u}\right)$, with the normalization factor $\gamma = \sqrt{N_{\rm r} N_{\rm t} / N_{cl} N_{ray}}$. For the uniform linear array (ULA), the antenna array responses at the transmitter/receiver, are represented by $\mathbf{a}_{\rm t}\left(\phi_{\rm t}^{l,u}\right) = \frac{1}{\sqrt{N_{\rm t}}}\left[1, e^{j2\pi\phi_{\rm t}^{l,u}}, ..., e^{j2\pi(N_{\rm t}-1)\phi_{\rm t}^{l,u}}\right]^H \in \mathbb{C}^{N_{\rm t} \times 1}$ and $\mathbf{a}_{\rm r}\left(\phi_{\rm r}^{l,u}\right) = \frac{1}{\sqrt{N_{\rm r}}}\left[1, e^{j2\pi\phi_{\rm r}^{l,u}}, ..., e^{j2\pi(N_{\rm r}-1)\phi_{\rm r}^{l,u}}\right]^H \in \mathbb{C}^{N_{\rm r} \times 1}$, respectively. It is important to note that the THz channel $\mathbf{H}$ in the spatial domain is effectively transformed into the equivalent channel in the beamspace domain $\mathbf{H}_b$, based on the DFT operations in lens antenna array (see [10] for details).

C. Problem Statement

In the considered hybrid analog-digital beamspace massive MIMO system, we focus on achieving analog beam selection for the transmitter and the receiver $\mathbf{S}_{\rm t}$ and $\mathbf{S}_{\rm r}$, under the assumption of known precoding/combining matrices and known beamspace channel. This problem can be formally stated as [12]

$$\begin{aligned} &\min_{\mathbf{S}_{\rm t}, \mathbf{S}_{\rm r}} \ ||\mathbf{H}_b - \mathbf{S}_{\rm r} \mathbf{W}_{\rm BB} \mathbf{F}_{\rm BB}^H \mathbf{S}_{\rm t}^H||^2 \\ &s.t. \\ &\mathbf{S}_{\rm r} \in \mathcal{S}_{\rm r}, \\ &\mathbf{S}_{\rm t} \in \mathcal{S}_{\rm t}, \end{aligned} \tag{1}$$

where $\mathcal{S}_{\rm t}$ and $\mathcal{S}_{\rm r}$ are the analog beam selection candidate sets at the transmitter and the receiver, respectively. The optimal solution for acquiring the analog beam selection variables $\mathbf{S}_{\rm r}$ and $\mathbf{S}_{\rm t}$ can be obtained by exhaustive search method, which is computationally expensive and practically infeasible for a beamspace massive MIMO system.

III. Solution Approach

In this section, the training sample set acquisition, the Swish-driven GoogleNet, the transfer learning and the ensemble learning are subsequently elaborated as our solution approach to (1).

A. Sample Set Acquisition

We consider the network parameters of path gain, transmit power, AoA, and AoD constituting $4N_{cl} \times N_{ray} + 2$ random real-valued features with one feature for the transmit power of the transmitter, one feature for the path gain, $2N_{cl} \times N_{ray}$ features for the AoDs/AoAs of the transmitter/receiver, and also $2N_{cl} \times N_{ray}$ features for the real and imaginary parts of the complex-valued gain to form a data sample. In the following, we conduct a normalization process, a Gaussian mixture model (GMM) fitting, and a labeling operation over the samples.

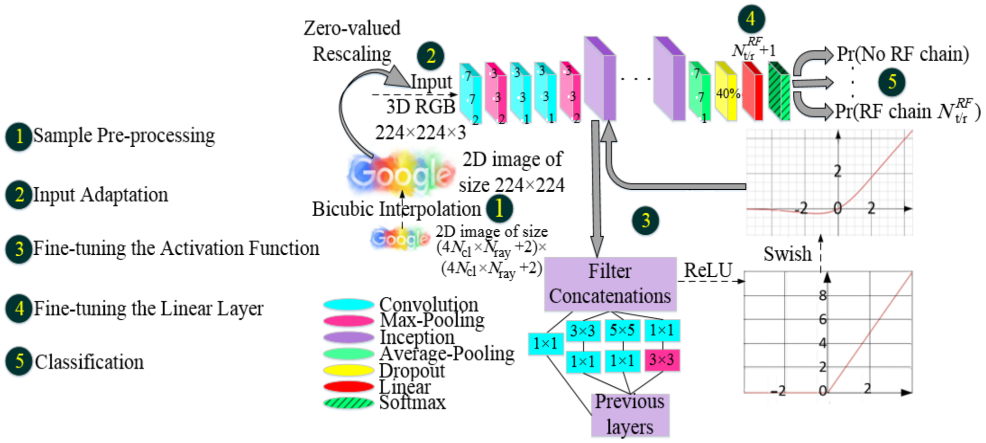


Fig. 2: Architecture of GoogleNet, modifications performed on training samples to fit into the input layer, replacing ReLU with Swish, and setting the number of linear layer classes from 1000 to N_{t}^{RF}+1 (or N_{r}^{RF}+1).

1) Normalization: Due to the diversity in sample ranges (e.g., the transmit power is based on dB, while the AoDs are within [0,2π]), a normalization pre-processing needs to be accomplished for each feature of samples as $\bar{a}_f^m = \left[a_f^m - \text{Mean}(a_f^m)\right] \times \left[a_f^{\text{max}} - a_f^{\text{min}}\right]^{-1}$, where a_f^m indicates the value of the fth feature in the mth sample and Mean(a_f^m) is the mean of all a_f^m. In addition, a_f^{max} and a_f^{min} denote the maximum and the minimum values of the fth feature among the entire samples, respectively. Hence, the mth sample as a feature row vector can be characterized as $\mathrm{z}_m \in \mathbb{C}^{1\times(4N_{cl}\times N_{ray}+2)}$ with $4N_{cl} \times N_{ray} + 2$ normalized features.

2) GMM Fitting: Since the beamspace channel features ϕ_t, ϕ_r, and α follow a Gaussian distribution [13], we adopt a GMM for appropriately fitting the beamspace channel. In doing so, we have $\tilde{\mathbf{H}}_b = A \times \left(\sum_{k=1}^{K} w_k \text{exp}\left(-\frac{(\phi_\text{r}-\mu_{\phi_{\text{r}_k}})^2}{2\sigma^2_{\phi_{\text{r}_k}}} - \frac{(\phi_\text{t}-\mu_{\phi_{\text{t}_k}})^2}{2\sigma^2_{\phi_{\text{t}_k}}} - \frac{(\phi_\text{r}-\mu_{\alpha_k})^2}{2\sigma^2_{\alpha_k}}\right)\right)$, with the GMM-fitted beamspace channel $\tilde{\mathbf{H}}_b$, the GMM amplitude A, and K Gaussian components, where $w_k \in [0,1]$ is the weight of the Gaussian component k and $\sum_{k=1}^{K} w_k = 1$. Note that in $\tilde{\mathbf{H}}_b$, the central coordinates are $(\mu_{\phi_{\text{r}_k}}, \mu_{\phi_{\text{t}_k}}, \mu_{\alpha_k})$, whereas $\sigma_{\phi_{\text{r}_k}}$, $\sigma_{\phi_{\text{t}_k}}$, and σ_{α_k} indicate their corresponding standard deviation. In vector representation, the Gaussian component k can be expressed as $q_k = [w_k, \mu_{\phi_{\text{r}_k}}, \mu_{\phi_{\text{t}_k}}, \mu_{\alpha_k}, \sigma_{\phi_{\text{r}_k}}, \sigma_{\phi_{\text{t}_k}}, \sigma_{\alpha_k}]$. Equivalently, the spatial features of the samples based on all of the Gaussian components can be given by $\mathbf{q} = [A; q_1; q_2; ...; q_K]^T = [A, \mu_{\phi_{\text{r}_1}}, \mu_{\phi_{\text{t}_1}}, \mu_{\alpha_1}, \sigma_{\phi_{\text{r}_1}}, \sigma_{\phi_{\text{t}_1}}, \sigma_{\alpha_1}, \mu_{\phi_{\text{r}_2}}, \mu_{\phi_{\text{t}_2}}, \mu_{\alpha_2}, \sigma_{\phi_{\text{r}_2}}, \sigma_{\phi_{\text{t}_2}}, \sigma_{\alpha_2}, ..., \mu_{\phi_{\text{r}_K}}, \mu_{\phi_{\text{t}_K}}, \mu_{\alpha_K}, \sigma_{\phi_{\text{r}_K}}, \sigma_{\phi_{\text{t}_K}}, \sigma_{\alpha_K}]^T$. Finally, the optimal vector $\mathbf{q}$, which is used to model the beamspace channel distribution can be determined according to [14].

3) Labeling: The cost function for evaluating the analog beam selection decisions (i.e., labeling) is the objective in (1), which equivalently optimizes the achievable SE [12]. The labeling phase is a multi-class mapping operation that determines the optimum beam and its corresponding RF candidates obtained from [15], wherein each RF chain is a class label which analog beams are assigned to.

B. GoogleNet Architecture

As an off-the-shelf pre-trained network, GoogleNet has been trained by the well-known datasets (e.g., ImageNet) beforehand, while its weights, biases, and other training parameters have already been set. According to Fig. 4, the network has 22 layers with an input layer of size 224×224×3 for receiving a two-dimensional (2D) image of width and length 224 and 3 channels of RGB (i.e., red, green, and blue). The main parts in GoogleNet architecture are its inception modules that incorporate multiple convolutions, kernels, and max-pooling layers, simultaneously within a single layer. The main activation function in GoogleNet is ReLU, which is computationally cheap and embedded into a filter concatenation layer within the inception module (see Fig. 4) for improved training performance. By going deeper in GoogleNet architecture as observed in Fig. 4, the linear layer of size 1000 is followed by a dropout layer with 40% ratio of dropped outputs and connected to a Softmax activation function with 1000 classes.

C. Swish-driven GoogleNet

Despite its accurate classification capability, the performance of GoogleNet can still be improved by minor architectural modifications. For instance, the authors in [16] proposed to substitute the ReLU activation functions in GoogleNet with the Leaky-ReLU (an extension of the conventional ReLU) for faster convergence. In [17], the large convolutional filters in GoogleNet were factorized into smaller ones, and this modification benefited for the middle layers of GoogleNet. In this paper, we modify the ReLU activation functions in the filter concatenation layer of the inception modules (see Fig. 4) in the GoogleNet architecture by the Swish [8]. The latter is a self-gated, smooth, and non-monotonic activation function recently proposed by the Google Brain Team. By definition, the Swish activation function for an any input x can be given by $f^{\text{Swish}}(x) = x.f^{\text{Sigmoid}}(x) = \frac{x}{1+e^{-x}}$. The numerical results in [8] indicate that the Swish is more accurate than the ReLU (and its alternative extensions, such as Leaky-ReLU) with a similar level of computational complexity, especially in deeper architectures.

D. Transfer Learning

To fit the size of the samples into the input layer of the fine-tuned Swish-driven GoogleNet, certain modifications need to be accomplished in accordance with Fig. 4. First, we extend the dimensionality of a typical sample z_m of size $(4N_{cl}\times N_{ray}+2)$ into a matrix form of size $(4N_{cl}\times N_{ray}+2)\times(4N_{cl}\times N_{ray}+2)$ as a 2D image. Then, we perform an image resizing through the interpolation technique to transform each sample onto the size of 224×224. Specifically, we use bicubic interpolation that can preserve the quality of the primary image by extracting the most determinant properties (which correspondingly are related to the most dominant features of the sample in our case). The 224×224 resized 2D image of z_m is eventually extended into a three dimensional (3D) image by using zero-valued rescaling. To do so, the RGB color triplet for each pixel is set to zero, thus leading to a 3D RGB image of size $224\times224\times3$ to feed the input layer of the GoogleNet.

We further fine-tune the final linear layer of the GoogleNet by setting N_{t}^{RF}+1 classes for the transmitter (or N_{r}^{RF}+1 for the receiver), which trains the GoogleNet to map any sample (beam) onto the correct class (RF chain). During the training process, the beamspace channel feature space is processed through the layers of the GoogleNet, while its main features (energy-focused features of the beam) are extracted. The Softmax classifier eventually learns a multi-class mapping based on the labeled samples obtained from [15]. The probability of the ith RF chain being selected by the Softmax function is $\delta(N_{\mathrm{t}}^{RF})_i=\left[e^{\left(N_{\mathrm{t}}^{RF}\right)_i}\right]\times\left[\sum_{i=1}^{|N_{\mathrm{t}}^{RF}|}e^{\left(N_{\mathrm{t}}^{RF}\right)_i}\right]^{-1}$.

Finally, as observed in Fig. 4, a modified version of the GoogleNet is trained by fine-tuning its linear layer and activation functions. This approach is known as transfer learning, whereby the main layers of a pre-trained network are directly imported into the new application, while other layers remain unchanged. By doing so, the fine-tuned GoogleNet learns analog beam selection at the transceivers based on the beamspace channel feature space, while its internal weights, biases, and other parameters are mostly fixed.

E. Enhancing Accuracy via Ensemble Learning

We further improve the accuracy of the proposed procedure for analog beam selection through the ensemble learning technique. By doing so, we train a strong ensembled model that combines the predictions made by distinct weak learners (e.g., the Swish-driven GoogleNet modules in this paper) to achieve a more precise model. To do so, a gradient boosting (gradBoost) mechanism [9] is adopted, wherein we sequentially train the weak learners.

In order to form an ensembled model as in Fig. 3, we adopt M_1 random subsets $Z_m(m\in M_1)$ of the entire training sample set $\mathcal{Z}$, where the weak learners are trained over different subsets. For any sample $z_m\in Z_m$ of size $\mathbb{C}^{1\times(4N_{cl}\times N_{ray}+2)}$, the weak learner performs a classification and assigns a specific class from $\omega_m\in\Omega=\{0,...,N_{\mathrm{t/r}}^{RF}\}$.

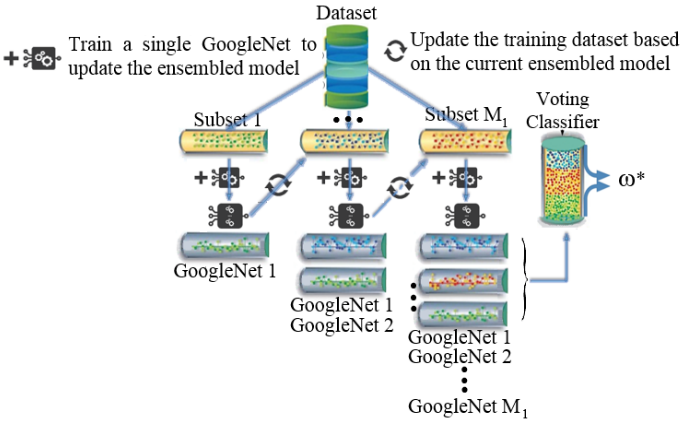

Dataset
Train a single GoogleNet to update the ensembled model
Update the training dataset based on the current ensembled model
Subset 1
Subset M1
Voting Classifier
ω*
GoogleNet 1
GoogleNet 1
GoogleNet 2
GoogleNet 1
GoogleNet 2
GoogleNet M1

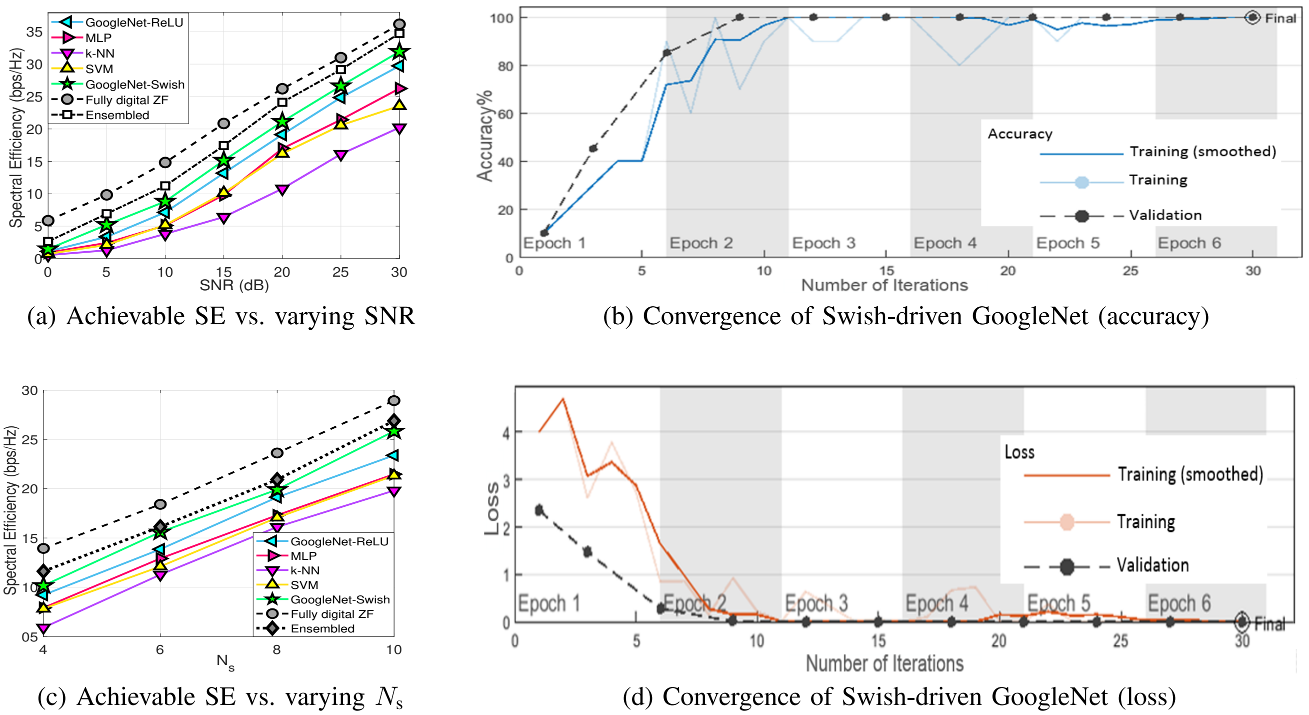
Fig. 3: Ensemble learning schematic.

The goal in each step is boosting the training accuracy of the current weak learner through focusing on the misclassified observations made by the previous ones. The misclassified samples are injected forward to train the next weak learner more efficiently. The strong ensembled learner then adopts a majority voting mechanism based on a weighted summation of M_1 weak learners. To this aim, a voting counter $\Psi(\omega)\in\mathbb{N}^{1\times\Omega}$ indicates the number of classifiers, which adopted the RF chain class ω. The weighted summation is given by $\Phi_{M_1}^{ens}=\sum_{m=1}^{M_1}c_m\Psi_m(\omega)$, where c_m denotes the weight (impact) of the mth Swish-driven GoogleNet. Indeed, the more accurate a weak learner is, the more it contributes to the strong ensembled model. The strong ensembled learner is therefore generally less biased than the weak learners, since the misclassified observations are efficiently propagated and learned along the ensembling chain. The challenge here is to select the optimal order of the classifiers to be trained within the ensembling chain, i.e., obtaining the optimal order of $\Phi_{M_1}^{ens}$ is complicated, especially for a long ensembling chain.

Instead of optimizing the said order globally, we are seeking for the best possible pairs of $(c_m,\Psi_m(\omega))$ to be built locally and added iteratively in a sub-optimal approach. The strong ensembled model can be recurrently formulated as $\Phi_m^{ens}=\Phi_{m-1}^{ens}-c_m\nabla_{\Phi_{m-1}^{ens}}E(\Phi_{m-1}^{ens})$, whereby the best possible pair $(c_m,\Psi_m(\omega))$ can be obtained as $(c_m,\Psi_m(\omega))=\arg\min_{c,\Psi(\omega)}E(\Phi_{m-1}^{ens}+c\Psi(\omega))$, with $E(.)$ denoting the strong ensembled learner fitting error. Finally, the RF chain class ω, which maximizes the voting counter $\Psi_m(\omega)$ by contributing M_1 weak learners and their impacts, is adopted by the strong ensembled learner as $\omega^*=\arg\max_{\omega\in\Omega}\frac{1}{M_1}\sum_{m=1}^{M_1}c_m\Psi_m(\omega)$.

IV. Simulation Results

We consider a clustered THz channel with $N_{cl}=4$ clusters and $N_{ray}=2$ propagation rays in each cluster. The signal wavelength is $\lambda=1.36$, the AoAs and the AoDs are uniformly distributed within $[-\frac{1}{2},\frac{1}{2}]$, while the complex-valued gain follows $\mathcal{CN}(0,\ 1)$. Simulations are performed for a lens-aided MIMO system equipped with $N_{\mathrm{r}}=64$, $N_{\mathrm{t}}=256$, and $N_{\mathrm{r}}^{\mathrm{RF}}=N_{\mathrm{t}}^{\mathrm{RF}}=4$. For the simulations related to the GoogleNet



(a) Achievable SE vs. varying SNR

(b) Convergence of Swish-driven GoogleNet (accuracy)

(c) Achievable SE vs. varying N_s

(d) Convergence of Swish-driven GoogleNet (loss)

Fig. 4: Convergence and performance of fine-tuned GoogleNet for analog beam selection.

TABLE I:
GoogleNet configuration

Parameter	Value
TrainingSize	70%
ValidationSize	30%
MiniBatchSize	128
MaxEpochs	6
Shuffle	every epoch
InitialLearnRate	1e-3
ValidationFrequency	3

as indicated in Table I, we used 70% of the sampling data for the training and the rest were utilized for the validation. Moreover, the "MiniBatchSize" shows the number of images used at each iteration of training/validation. The maximum number of training epochs is indicated by "MaxEpochs" and the "Shuffle" field is added every epoch, which randomly initiates a new datastore with the same training/validation data. The initial learning rate "InitialLearnRate" slows down the learning process in the transferred layers owing to its adopted small value and the "ValidationFrequency" field specifies that the validation is performed every three iterations during training. The achievable SE of a hybrid analog-digital beamspace system can be expressed as $SE = \log_2 \big| \mathbf{I}_{N_s} + \frac{\rho}{\sigma^2 N_s} R_n^{-1} (\mathbf{W}_{\rm BB})^H (\mathbf{S}_{\rm r})^H \mathbf{H}_b \mathbf{S}_{\rm t} \mathbf{F}_{\rm BB} (\mathbf{F}_{\rm BB})^H (\mathbf{S}_{\rm t})^H (\mathbf{H}_b)^H \mathbf{S}_{\rm r} \mathbf{W}_{\rm BB} \big|$, where ρ indicates the average received power and $R_n = (\mathbf{W}_{\rm BB})^H (\mathbf{S}_{\rm r})^H \mathbf{S}_{\rm r} \mathbf{W}_{\rm BB}$ is the noise covariance matrix after combining.

The analog beam selection baseline strategies MLP, k-NN, and SVM with the same internal configurations as in [3], the conventional ReLU-driven GoogleNet, the modified Swish-driven GoogleNet, and the ensemble learning schemes are investigated for comparison in terms of the achievable SE. Additionally, the fully digital zero-forcing (ZF) strategy by using all of the beams at the transceivers is the optimal benchmark baseline.

First, we assess the convergence accuracy and loss ratios for the training/validation process of the proposed Swish-driven GoogleNet scheme in Figs. 4(b) and 4(d), respectively. Clearly, the training/validation process is inaccurate over the first iterations. That is because the weights and biases of the input layer and the linear layer are not well fine-tuned with the sampling data. Gradually, as the iterations progress, the training/validation accuracy improves (tends to 100%), while the training/validation loss declines (tends to 0).

Further, we analyze the performance of our proposed schemes in a comparative fashion. The benchmark fully-digital ZF strategy with $N_{\rm t}^{RF} = 256$ and $N_{\rm r}^{RF} = 16$ RF chains apparently has the largest achievable SE in Fig. 4(a) and Fig. 4(c) at the expense of severe system complexity, energy consumption, and hardware cost. Fig. 4(a) with varying SNR in 0dB$\sim$30dB and $N_{\rm t}^{RF} = N_{\rm r}^{RF} = N_{\rm s}$, where $N_{\rm s} = 4$, indicates that by increasing the SNR the achievable SE improves for all the baselines. According to Fig. 4(c) with varying $N_{\rm s}$ in 4$\sim$10, where $N_{\rm t}^{RF} = N_{\rm r}^{RF} = N_{\rm s}$ and SNR = 10dB, the achievable SE improves for a higher number of simultaneous data streams. Our proposed ensemble learning scheme is the most superior over the other baselines and is the closest scheme to the benchmark due to its better accuracy. This scheme according to Fig. 4(a) improves the achievable SE of the MLP scheme [3] at SNR = 30dB by up to 27%. Similarly,

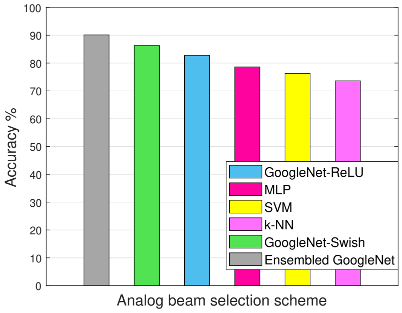


Fig. 5: Aanalog beam selection accuracy comparison.

TABLE II:
GoogleNet-based analog beam selection accuracy comparison.

Architecture/Function	RMSPROP	ADAM	SGDM
GoogleNet-ReLU	83.4%	81.37%	82.22%
GoogleNet-Swish	86.21%	85.27%	86.93%

at SNR = 30dB, the proposed Swish-enabled GoogleNet and the conventional ReLU-driven GoogleNet schemes achieve better performance than other strategies MLP, SVM, and k-NN, by exhibiting 18% and 12% achievable SE gain compared to the MLP scheme [3], respectively.

In Fig. 5, under the same configurations in Fig. 4(c) with $N_\mathrm{s} = 4$, the ensemble learning strategy, the Swish-driven GoogleNet and the conventional ReLU-driven GoogleNet with 90%, 86% and 83% accuracy average accuracy are the most accurate schemes. The reason is that retraining/modifying the pre-trained networks such as GoogleNet based on transfer learning for the classification tasks (e.g., analog beam selection) is more accurate than training a deep network such as MLP [3] from scratch. Using the transfer learning method, the parameters in a pre-trained deep structure are mostly kept unchanged, while a few certain parameters are fine-tuned based on samples. We further examine the accuracy of the conventional ReLU-driven GoogleNet as well as the fine-tuned Swish-driven GoogleNet schemes by applying different training schemes e.g., root mean square propagation (RMSPROP), adaptive moment estimation (ADAM), and stochastic gradient descent method (SGDM), as demonstrated in Table II. One can observe that the Swish-driven GoogleNet scheme trained by the SGDM can achieve the best analog beam selection accuracy. Note that unlike the training accuracy (in the presence of labeled samples) or the validation accuracy (with limited number of non-labeled samples) in Fig. 4(b) that both approach 100%, the accuracy in Fig. 5 is obviously lesser owing to operating on large non-labeled analog beam selection samples.

V. Conclusions

In this paper, we proposed a novel deep learning technique to address the analog beam selection problem in a THz beamspace MIMO system. Specifically, we retrained the pre-trained off-the-shelf GoogleNet for learning the analog beam selection based on the concept of transfer learning. Then, we fine-tuned the GoogleNet by enabling the Swish activation function for better analog beam selection precision. Finally, an ensemble learning technique employed for boosting the precision beyond that of the conventional fine-tuned GoogleNet. Simulations revealed a remarkable enhancement in accuracy as well as in the achievable SE.

Acknowledgement

This work is supported by the Academy of Finland: (a) ee-IoT n.319009, (b) EnergyNet n.321265/n.328869, and (c) FIREMAN n.326270/CHISTERA-17-BDSI-003; and by JAES Foundation via STREAM project.

References

[1] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis and measurements," *IEEE Trans. Antennas Propag.,* vol. 61, no. 7, pp. 3814-3827, Jul. 2013.

[2] I. Orikumhi, J. Kang, H. Jwa, J. H. Na and S. Kim, "SINR maximization beam selection for millimeter-wave beamspace MIMO systems," *IEEE Access,* vol. 8, pp. 185688-185697, 2020.

[3] C. Anton-Haro and X. Mestre, "Learning and data-driven beam selection for Millimeter-Wave communications: an angle of arrival-based approach," *IEEE Access*, vol. 7, pp. 20404-20415, 2019.

[4] X. Ma, Z. Chen, Z. Li, W. Chen and K. Liu, "Low complexity beam selection scheme for terahertz systems: A machine learning approach," *IEEE ICC Workshops*, Shanghai, China, pp. 1-6, 2019.

[5] A. Klautau, P. Batista, N. Gonzalez-Prelcic, Y. Wang and R. W. Heath, "5G MIMO data for machine learning: application to beam selection using deep learning," *Proc., ITA*, pp. 1-9, 2018.

[6] C. Szegedy et al., "Going deeper with convolutions," *IEEE Conf. Comp. Vis. Patt. Recogn. (CVPR)*, Boston, MA, 2015, pp. 1-9, 2015.

[7] L. Y. Pratt and T. Sebastian, "Machine learning," *Special issue on inductive transfer,* July, 1997.

[8] P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: A selfgated activation function." *arXiv preprint, arXiv:1710.05941*, Oct. 2017.

[9] T. Hastie, R. Tibshirani, J. H. Friedman, "10. Boosting and additive trees", *The Elements of Statistical Learning (2nd ed.), Springer*, pp. 337–384, Nov., 2009.

[10] W. Shen, X. Bu, X. Gao, C. Xing and L. Hanzo, "Beamspace precoding and beam selection for wideband millimeter-wave MIMO relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6301-6313, Dec. 2019.

[11] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128-137, Feb. 1987.

[12] M. Wang, F. Gao, S. Jin and H. Lin, "An overview of enhanced massive MIMO with array signal processing techniques," *IEEE J. Sel. Top. Signal. Process.*, vol. 13, no. 5, pp. 886-901, Sept. 2019.

[13] X. Wei, C. Hu and L. Dai, "Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 182-193, Jan. 2021.

[14] G. Celeux, S. Chretien, F. Forbes, and A. Mkhadri, "A component-wise EM algorithm for mixtures," *Journal of Computational and Graphical Statistics*. no.4 pp. 697—712, Jan., 2012.

[15] P. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212-2222, Jun., 2015.

[16] L. Balagourouchetty, J. K. Pragatheeswaran, B. Pottakkat and G. Ramkumar, "GoogLeNet-based ensemble FCNet classifier for focal liver lesion diagnosis," *IEEE J. Bio. Hlth. Inf.*, vol. 24, no. 6, pp. 1686-1694, June 2020.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," *IEEE Conf. Comput. Vis. Patt. Rec. (CVPR)*, Las Vegas, NV, 2016, pp. 2818-2826.