MMSENSE: DETECTING CONCEALED WEAPONS WITH A MINIATURE RADAR SENSOR

Kevin Mitchell [†] [§] [*] *Khaled Kassem* [†] [§] *Chaitanya Kaul* [‡] [§]
Valentin Kapitany [†] Philip Binner [†] Andrew Ramsay [‡]
Daniele Faccio [†] Roderick Murray-Smith [‡]

[†]School of Physics and Astronomy, University of Glasgow, UK, G12 8SU
[‡] School of Computing Science, University of Glasgow, UK, G12 8RZ

ABSTRACT

For widespread adoption, public security and surveillance systems must be accurate, portable, compact, and real-time, without impeding the privacy of the individuals being observed. Current systems broadly fall into two categories – image-based which are accurate, but lack privacy, and RF signal-based, which preserve privacy but lack portability, compactness and accuracy. Our paper proposes *mmSense*, an end-to-end portable miniaturised real-time system that can accurately detect the presence of concealed metallic objects on persons in a discrete, privacy-preserving modality. mmSense features millimeter wave radar technology, provided by Google's Soli sensor for its data acquisition, and TransDope, our real-time neural network, capable of processing a single radar data frame in 19 ms. mmSense achieves high recognition rates on a diverse set of challenging scenes while running on standard laptop hardware, demonstrating a significant advancement towards creating portable, cost-effective real-time radar based surveillance systems.

Index Terms— Real-time signal processing, mmWave radars, Vision Transformer

1. INTRODUCTION

Radar solutions developed on Frequency Modulated Continuous Wave (FMCW) technology have shown promising success through their ability to serve as a capable and versatile basis for computational sensing and short range wireless communication systems [1]. Such radars, operating at millimeter wave (mmWave) frequency, can be used for robust gesture generation and recognition [2, 3], and even measure distances with *mm* accuracy [4]. Furthermore, mmWave radars have the potential to serve as a basis for concealed metallic object detection (e.g. knives, guns etc) which presents a novel and most importantly, privacy-preserving manner of real-time surveillance. The principles of mmWave metal detection rely on the underlying physics of RF waves- radio frequency (RF) waves that fall in the 30–300 GHz range between microwaves and terahertz waves. This frequency band corresponds to wavelengths of 1–10 mm. Within various forms of spectral imaging (e.g. IR, UV), one chooses the waveband that interacts with the object/scene to be imaged, whilst ignoring any obfuscating features. The same is true when detecting metals concealed on humans, where the mmWave waveband is appropriate because the waves from a mmWave RF source pass through the thin layers of clothes, and are reflected highly by the body, plus any hidden objects between the two [5]. Figure 1 depicts this concept empirically. We believe there is a niche in the security field for a portable technology that can screen for illegal metallic weapons, whilst still allowing people to maintain their freedom and privacy in public without security bottlenecks and conventional image-capturing cameras.

Fig. 1: The RF waves from a Google Soli reflect from a person standing 1.5m away from the radar with and without a knife. To mitigate the variations in specular reflection, the subject rotated through 90° and an average over 60s was used in each scene. The reflected signal received via receiver 2 (of 3) of the Soli is converted into a range profile by computing an FFT over it, and plotted here. The difference between the two range profiles shows the potential of using mmWave radars like the soli, for detecting metallic objects.

This work is not the first to propose mmWave sensing for metal detection. Fusion of mmWave radar and vision data has already helped create accurate and effective object detection systems for autonomous driving applications [6]. The

[*]We acknowledge funding from the QuantIC Project funded by the EPSRC Quantum Technology Programme (grant EP/MO1326X/1), and Google. R.M.S acknowledges EPSRC grant EP/R018634/1, *Closed-loop Data Science*

[§]Equal Contribution

comparison in performance of Convolutional and Recurrent Neural Networks (CNNs and RNNs) for metal detection using magnetic impedance sensors has been extensively evaluated in [7]. Their system is compact, however they need to scan the complete scene, and restrict themselves to large metal sheets that are visible in the Line-of-Sight (LOS) of their sensor. [8] use mmWave radar sensors to alert robots to the presence of humans. Of most relevance to our study are [9, 10]- the former study, [9] provides a comprehensive guide on the metal detection capabilities of a $77 - 81$GHz radar chip but do so by comparing intensities of the reflected signal with the intensities of their own model of a body without the presence of the metallic object. Their work does not look at concealed metallic objects but ones that are already visible. [10] created an AI powered metal detection system capable of working in real-time. The system, however, processes their data differently, is prohibitively expensive, and is considerably bulkier than ours. One fundamental advantage our system has over all existing mmWave systems proposed for similar applications is the use of a radar sensor that has the widest Field of View (FOV), smallest form factor and least power consumption amongst its competitors. The Soli is capable of illuminating its surroundings with a 150^o FOV radar pulse. The lower power consumption of the Soli is due to the fact that it transmits 16 chirps at a pulse repetition frequency of 2000 Hz in a single burst (each burst is transmitted at 25Hz), and then stops transmitting until the next burst to save power [3]. This saves considerable power compared to mmWave radars used in existing works that continuously transmit chirps. Compared to current radar based surveillance systems, our technology does not need to sweep a scene to work, but provides inference on-the-fly by illuminating the scene with RF waves and processing the received signal.

Our work is intended to disrupt the trend of specialised surveillance and imaging systems which are becoming increasingly expensive to install and operate, by using an inexpensive, compact device capable of being mounted in various locations throughout open spaces, which can function in real time. To this end, we present the use of a commercial mmWave radar transceiver to detect the presence of concealed objects on people in real time, in a privacy preserving manner. We focus on high frequency (60GHz), short range (up to 3m) sensing using Google's Soli sensor, primarily due its miniature form factor, low power consumption, portability, and novel sensing characteristics. The Soli is designed for Human-Computer Interaction applications, and has shown success in 'macro' radar-based computational imaging tasks (detecting motion, gestures etc). Its application to detecting objects within the movement is unexplored and challenging. The Soli captures a superposition of reflected energy from different parts of a scene using high frequency RF waves: this results in poor lateral spatial resolution, while detecting even the smallest amount of motion. This makes metal detection challenging when there is plenty of movement in the scene i.e., in all practical real world scenarios. To mitigate this challenge, we propose a novel, real-time Vision Transformer model that can exploit semantic sequential relations in the preprocessed radar data and recognize the presence of a concealed metallic object on a person in a scene while ignoring objects such as wallets, keys, belts and mobile phones.

The following are our main contributions: (1) We present `mmSense` - a novel, end-to-end framework capable of detecting concealed metallic objects on people in a scene using only raw radar data without any human intervention. (2) `mmSense` is real-time, and can potentially use off-the-shelf commercially available hardware making it easy to replicate and deploy. (3) We open source *mmSense* including all our datasets and models with the hopes of facilitating further research in this novel field of Artificial Intelligence powered concealed metal detection with mmWave RF signals.

2. MMSENSE

Our mmSense pipeline comprises of three components – a Google Soli radar for data acquisition, an Intel RealSense D435 Time-of-Flight (TOF) camera for visualizing results, and an API capable of acquiring and processing radar data streams in real-time. A single burst of the Soli's transmitted signal is received across 3 antennas. For each RF illumination of the scene by the radar, it receives back a signal $I \in \mathbb{R}^{P \times C}$ where I is the imaged scene as perceived by the radar and P is the number of chirps received by the radar across its $C(=3)$ antennas. We operate the Soli at a frequency of 60GHz with 1GHz, the maximum permitted bandwidth BW. This gives us a range resolution $R_r = \frac{c}{2BW} = 15$cm, where c refers to the speed of light. This is the minimum distance that the radar can separate in its LOS between two distinct points. The Soli transmits and receives a series of chirps that are grouped into bursts- we define the number of chirps for our system to be 16, and collect bursts at 25Hz, giving us 25 bursts of Soli data in one second. In this configuration, we can detect up to a maximum range of 9.6m.

The Soli hardware has a corresponding API provided by Google and implemented in C++. We built a C++ application around this API which allows us to interface with the Soli radar in real-time (e.g. selecting a range profile) and receiving the bursts generated from the radar. Our application supports streaming the Soli bursts directly to a Python script using the ZeroMQ[1] messaging framework. The bursts are relayed immediately upon being received by the device with no additional buffering, and are ultimately parsed by a Python module which extracts both the parameters associated with the burst (e.g. a hardware-supplied timestamp) and the raw radar data.

After parsing the raw chirps and their timestamps, we create a Range Doppler [11] transformation of the signal. This is done via a series of Fast Fourier Transforms (FFT) applied to the data. First, we calculate the complex value range profile (RP) of the radar signal. This is done via an FFT of the radar chirps received by the 3 antennas. As the Soli's signal is a superposition of reflections from a scene, the RP data can be interpreted as how well the separate contributions of the RF scatters in the scene are resolved. This gives us an estimate of the geometry of the scene. A Complex Range Doppler (CRD) plot is then calculated as an FFT over each channel of the radar's complex value range profile. Here, the range represents the distance of the object in a scene from the Soli, and the Doppler corresponds to the radial velocity of the object towards the Soli. We use the magnitude of the CRD for our experiments, which is obtained in the following way, $\texttt{ARD}(r, d) = |\texttt{CRD}(r, d)|$, where `ARD` refers to the Absolute Range Doppler, r and d are the range and doppler bins, and

[1] `https://github.com/zeromq/cppzmq`

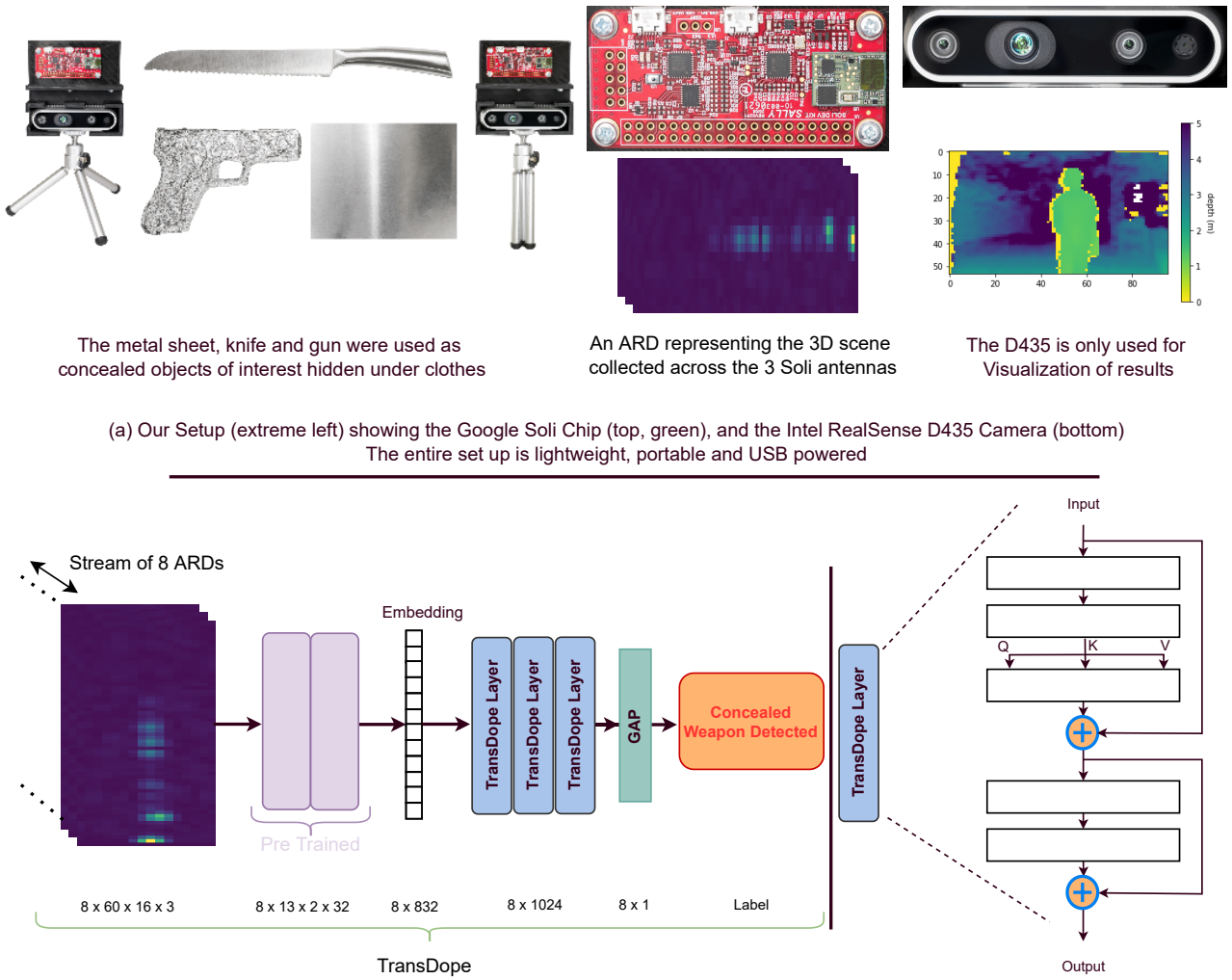

Fig. 3: TransDope (shown here) processes the radar data stream while preserving time-dependent information throughout.

$|\cdot|$ is the absolute value.

The ARD plots are processed using our novel Deep Neural Network, TransDope (Doppler Transformer, figure 3), capable of accurate real time classification of the input data stream. The input is a sequence of 8 ARD frames. TransDope contains two Time Convolution layers pretraind on a large collected dataset of ARD plots, an embedding layer to create a transformer compatible embedding of the convolution features, and transformer layers to learn long range time-dependent semantic features in the data. We first collect a large dataset of ARDs from various scenes with and without concealed metallic objects on actors. We then train a model with two Time Convolution and Max Pooling layers, the output of which is flattened and fed to a classification layer.

Following training, we discard the output layer, and use the two time convolution layers with the pre-trained weights as an initialization. Unlike standard convolutions that apply a weight kernel across all 8 ARDs concurrently, we apply convolutions sequentially to the 8 ARD frames to extract time-dependent features from them, and hence call them time convolutions. We then reshape the output of the last Max Pooling layer to create an embedding of the features. We also add positional encoding to each of the 8 ARD frames to preserve their sequence. Following this, we pass the embedding through 3 TransDope layers that extract semantic dependencies within the ARD's feature representation. These layers are the same as ViTs [12] encoder layers with the exception of having a convolutional layer following the multi head attention layer, instead of the dense layers, to reduce parameter size. We use global average pooling to reduce the transformer layer's features to a vector, which are then passed into the output layer. Our Time Convolution layers have 32 filters and a kernel size of 3×3. Our transformer layer has an embedding size of 128 and uses 2 attention heads. TransDope contains 0.8 million parameters, and can process a single ARD frame in 19 milliseconds on an Intel i9 8-core CPU. We train our model in TensorFlow 2 for 50 epochs, with a batch size of 8, and a learning rate of $1e-2$ which inversely decays after every 10 epochs. During inference, we feed 1 batch of 8 ARD frames through the model to get a classification.

3. EXPERIMENTS

To test the accuracy and flexibility of our technique, we collected 6 different scenes with varying characteristics, as de-

P \ S	A	B	C	D	E	F
Object	M	K	K	G	G	G
People	1	5	1	2	1	1
Dist. (m)	2.0	2.0	2.0	2.85	1.5	2.0
Closed	✗	✗	✗	✗	✓	✓
Accessories	✓	✓	✗	✗	✗	✗
Acc. (%)	95.1	86.9	88.4	89.0	74.6	79.2

Table 1: Distinct properties (P) of the scenes (S) we collected, along with the accuracy of TransDope to identify the object for that scene. Here, K, G and M refer to the Knife, Gun and Metal Sheet respectively, people is the number of individuals in the scene and distance (in metres) is the maximum distance from the Soli that the individuals in the scene walk up to. Closed refers to a scene where the walls are close to the radar, and the RF signal can bounce off adjacent walls. Accessories are everyday items belts, wallets, keys and phones that the people in the scene carry.

picted in table 1. Data was acquired in 4 instances for each scene: 2 with a metallic object hidden on a person, and 2 without. Each acquisition contains approximately 1500 frames of data. Before training our machine learning model, we collected roughly 15,000 frames of Soli data equally split into the two classes in various scenes, to pre-train the TransDope time convolutions. For each scene, we then trained TransDope, to predict a binary class for each ARD.

We carefully curated different scenes to portray real world situations where our system can be deployed. Scene A was an initial proof of concept where we used the metal sheet as the hidden object to verify the capabilities of our system. We were able to predict the presence of the sheet with 95.1% accuracy. Scene B replicates a crowded expansive scenario, such as an airport terminal. Here we crowded the scene with 5 people walking up to 2m away in radius from the setup. Each person in the scene was carrying everyday objects such as phones, keys, wallets, and belts; only one of these individuals had a knife on their person. Even in such a challenging setting, our system detected the presence of the knife with up to 86.9% accuracy. In Scenes C and D, we observed the effects of changing the hidden object from a knife to a gun. This is important as different objects have different characteristic specular reflections. As seen from our results, the performance of our system held when switching the metallic object from the knife to the gun. Scenes A to D were all open scenes, i.e. the data was not acquired with constricting walls. This results in no multipath RF signals received by the Soli receiving antennas. In Scenes E and F, we tested the effects of keeping our set up in a closed setting and noticed performance decreased. Our results are summarized in table 1 and visualized in figure 4.

Ablations. Table 5 shows the effect of varying the amount of sequential information provided to TransDope, as well as varying the various blocks of TransDope. The experiments show that each individual component in our model contributes to performance boosts in terms of metal detection accuracy. We chose 8 ARD frames per sequence as the input to TransDope due to it providing the best accuracy versus execution time. Having multiple sequences of ARDs does further boost performance, but it also doubles (for 2 ARD sequences of 8 ARD frames - $2*8$), and quadruples (for $4*8$) the execution time for only minor gains in performance.

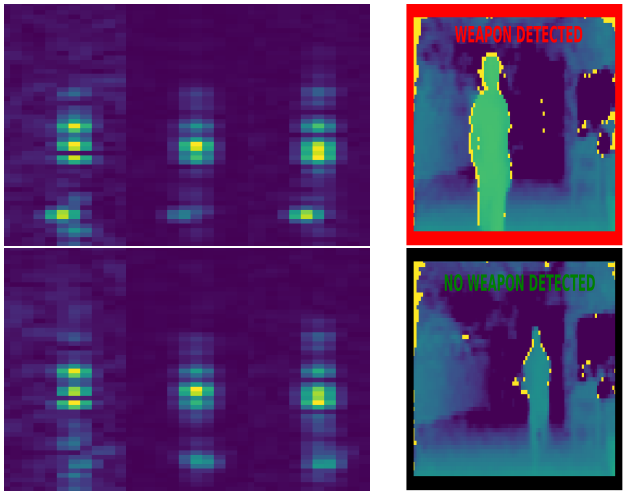

Fig. 4: Results for Scene C where the person has a knife hidden under a jacket. On the left are the ARDs for the three receiver antennas, and on the right is the output visualization. The red TOF output denotes the knife being detected, and the black is the misclassification due to the specular reflection from the knife not providing a strong signal.

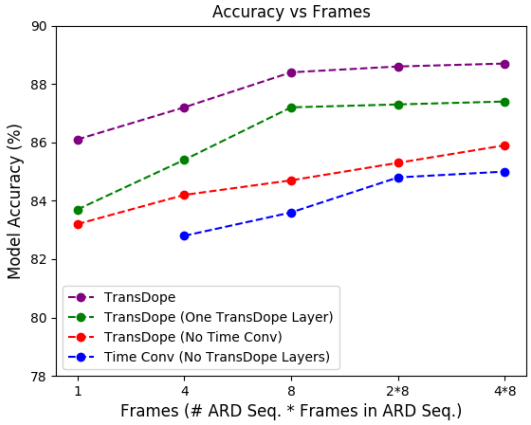

Fig. 5: Effects of varying the number of frame sequences and model layers on performance on Scene C.

4. CONCLUSIONS AND LIMITATIONS

Our paper proposed *mmSense*, an AI assisted concealed metal detection system whose technology is centred on an mmWave transceiver. *mmSense* can detect the presence of concealed weapons capable of mass harm even when scenes are crowded. It does so in real-time, and without compromising the privacy of the individuals in the scene. Our system however, has certain limitations in terms of performance degradation due to the effects of multipath RF waves at the receiver that occurs in close walled areas, as well as the inherent lack of spatial context from which commercial mmWave sensors suffer. We believe that investigating multi-modal sensor fusion of the radar and TOF data to add spatial awareness to the 'sensing' ability of *mmSense* may help to alleviate these drawbacks, and would be a fitting next step to extend this technology.

5. REFERENCES

[1] Ismail Nasr, Reinhard Jungmaier, Ashutosh Baheti, Dennis Noppeney, Jagjit S. Bal, Maciej Wojnowski, Emre Karagozler, Hakim Raja, Jaime Lien, Ivan Poupyrev, and Saverio Trotta, "A Highly Integrated 60 GHz 6-Channel Transceiver With Antenna in Package for Smart Sensing and Short-Range Communications," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 9, pp. 2066–2076, 2016.

[2] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges, "Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, New York, NY, USA, 2016, UIST '16, p. 851–860, Association for Computing Machinery.

[3] Eiji Hayashi, Jaime Lien, Nicholas Gillian, Leonardo Giusti, Dave Weber, Jin Yamanaka, Lauren Bedal, and Ivan Poupyrev, "RadarNet: Efficient Gesture Recognition Technique Utilizing a Miniature Radar Sensor," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2021, CHI '21, Association for Computing Machinery.

[4] Changzhan Gu and Jaime Lien, "A two-tone radar sensor for concurrent detection of absolute distance and relative movement for gesture sensing," *IEEE Sensors Letters*, vol. 1, no. 3, pp. 1–4, 2017.

[5] J. E. Bjarnason, T. L. J. Chan, A. W. M. Lee, M. A. Celis, and E. R. Brown, "Millimeter-wave, terahertz, and mid-infrared transmission through common clothing," *Applied Physics Letters*, vol. 85, no. 4, pp. 519–521, 7 2004.

[6] Zhiqing Wei, Fengkai Zhang, Shuo Chang, Yangyang Liu, Huici Wu, and Zhiyong Feng, "Mmwave radar and vision fusion for object detection in autonomous driving: A review," *Sensors*, vol. 22, no. 7, 2022.

[7] Sungjae Ha, Dongwoo Lee, Hoijun Kim, Soonchul Kwon, EungJo Kim, Junho Yang, and Seunghyun Lee, "Neural network for metal detection based on magnetic impedance sensor," *Sensors*, vol. 21, no. 13, 2021.

[8] Daniel Mitchell, Jamie Blanche, Sam T Harper, Theodore Lim, Valentin Robu, Ikuo Yamamoto, and David Flynn, "Millimeter-wave foresight sensing for safety and resilience in autonomous operations," *arXiv preprint arXiv:2203.12987*, 2022.

[9] Yixuan Lu, Weixi Chen, Haipeng Liu, and Anfu Zhou, "Study on feasibility of remote metal detection using millimeter wave radar for convenient and efficient security check," *CCF Transactions on Pervasive Computing and Interaction*, vol. 3, no. 3, pp. 284–299, 2021.

[10] David A. Andrews, Stuart William Harmer, Nicholas J. Bowring, Nacer D. Rezgui, and Matthew J. Southgate, "Active millimeter wave sensor for standoff concealed threat detection," *IEEE Sensors Journal*, vol. 13, no. 12, pp. 4948–4954, 2013.

[11] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev, "Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar," *ACM Trans. Graph.*, vol. 35, no. 4, jul 2016.

[12] Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai, "An image is worth 16×16 words: Transformers for image recognition at scale," 2021.