Real-Time Multi-Task Tactile Sensing Using Photoelastic Fringe Patterns and Deep Learning

Author: Yannick Serrien[1,*] and **Supervisors:** Georgy Filonenko, Gerwin Smit, Michaël Wiertlewski, Agnes Kereszturi, Giuseppe Vitrani[1]

[1]Delft University of Technology, The Netherlands

[*]Correspondence: yannick.serrien110@gmail.com

Abstract We present a low-cost, camera-based tactile sensor that leverages the photoelastic effect—interference fringes that appear under stress—to estimate contact force, position, and shape. Each fringe image is recorded at 50 Hz and processed by a multi-task neural network that predicts (i) the normal force (F_z), (ii) the 2D contact location (x, y), and (iii) the shape class of the object. Two sensor variants were developed: *Sensor 1*, a layered design with fewer visible fringes, and *Sensor 2*, an integrated structure with improved fringe clarity. Both were evaluated using a ResNet-18 and a lightweight custom CNN, under three augmentation pipelines: grayscale images with 10 noisy augmented samples each, RGB images with 3 noisy augmentations, and RGB images with 3 clean (noise-free) augmentations.

The base dataset includes nearly 15,000 synchronised samples of high-frequency fringe images and force signals. With augmentation, this was expanded to around 45,000 or 150,000 samples depending on the pipeline. The best results were achieved using Sensor 1 and ResNet-18 trained on grayscale images with 10 augmentations per input image. This configuration yielded a force MSE of 0.0213 N^2, a contact-point RMSE of 0.4462 mm, and 96.24% shape classification accuracy. Notably, even RGB images with only three augmentations per sample reached similar performance levels. These findings highlight that full-colour input and lightweight augmentation remain effective for accurate, scalable tactile sensing. Our modular learning pipeline generalises across sensor variants and data regimes, enabling robust, high-frequency tactile inference suitable for real-world deployment.

Keywords Photoelasticity · Tactile Sensing · Multi-task Learning · Machine Learning · Force Estimation · Contact Localisation · Shape Detection

1 Introduction

Robots perceive the world through sight and sound, but touch reveals how firmly they can push. Reliable tactile feedback is essential for delicate everyday tasks: holding a coffee mug without dropping it, enabling a prosthetic hand to pick up a berry without crushing it, or guiding a robot to clean dust from a glass table without leaving scratches. Traditional force sensors measure single points accurately but reveal little about how pressure spreads across surfaces, while cameras capture detailed images yet lack tactile sensitivity.

Currently available tactile sensors have significant limitations. Optical skins, such as GelSight, show that adding shape alone improves grasp success and slip detection [1], [2]. However, GelSight does not directly measure force values. On the other hand, capacitive or barometric arrays can measure force directly but provide only coarse spatial detail, missing precise shape and location information crucial for nuanced tasks. This lack of spatial and force precision becomes especially problematic in delicate applications—for example, grasping a strawberry off-centre may cause the gripper to slip or apply excessive force to a small region, damaging the fruit. In contrast, a sensor that can measure both force magnitude and its precise distribution would allow the system to grasp the strawberry closer to its centre, applying just enough force to hold it securely without crushing it.

Photoelasticity bridges these limitations. A transparent, stress-sensitive polymer generates interference fringes—*isochromatics*—when loaded. These fringes simultaneously encode force magnitude, direction, and distribution. Previous photoelastic sen-

sors demonstrated a force resolution of 0.5N for forces up to 8N but lacked localisation of the contact point. [3] Polyurethane-sheet sensors could measure tangential forces, but ignored shape recognition. [4] A photoelastic tactile sensor, capable of simultaneously determining multi-axis forces, precise contact locations, and object shapes in real time, would significantly advance robotic tactile sensing.

In this work, we propose a camera-based tactile sensor system that reads photoelastic fringe patterns and predicts three tactile outputs in parallel: the normal force (F_z), the 2D contact point (x, y), and the indentor shape. Our system is designed to be low-cost ($\approx$ 200 euros), scalable, and data-driven. It is built around a modular architecture that supports both grayscale and RGB imaging pipelines, as well as various augmentation strategies designed to test performance under noisy, realistic, and ideal conditions.

Research Question:
How effectively can a camera-based photoelastic tactile sensor, combined with a lightweight multi-task CNN, estimate real-time tactile information, and how do specific sensor design and training choices influence its performance?

Sub-questions:

1. How does changing the internal structure (layered versus integrated design) affect sensing accuracy?
2. What impact does fringe image quality (sharp versus blurred images) have on the network's predictive performance?
3. Which augmentation strategies (grayscale vs. RGB, noisy vs. clean) offer the best trade-off between accuracy, realism, and computational cost?

To investigate these questions, we developed two sensors: *Sensor 1*, a layered structure that produces lower-contrast fringes, and *Sensor 2*, an integrated design with improved fringe clarity. We tested both using two neural networks: a ResNet-18 and a custom lightweight CNN. Each network receives a single image as input and predicts all three tactile outputs simultaneously. To build a diverse training dataset, we collected nearly 15,000 synchronised samples using a 50 Hz camera and a 5 kHz force sensor. Depending on the augmentation pipeline—either 3 or 10 augmented versions per sample—this dataset was expanded to around 45,000 or 150,000 images. We experimented with grayscale inputs as well as full RGB images, with and without added noise and jitter, to examine how different representations and conditions affect learning.

All image preprocessing was designed to preserve the physical validity of the data. Fringe images were cropped so the sensor was centred, allowing for geometric transformations such as rotation. This ensured that spatial labels (force, location, shape) remained accurate. The final image size was adapted to each sensor's reflective region: 980×980 pixels for Sensor 1, and 1100×1100 pixels for Sensor 2.

The paper is structured as follows: section 2 reviews key developments in photoelastic sensing across both tactile and non-tactile domains, highlighting the need for simple, multimodal systems capable of full-field inference. section 3 details the design of the two sensor variants, the data acquisition setup, the augmentation pipeline, and the neural network architectures used. section 4 presents both quantitative and qualitative findings, analysing how sensor design, image quality, and augmentation strategies affect prediction performance across force, contact location, and shape classification tasks. This section also introduces the evaluation metrics used for assessing model accuracy. In section 5, we benchmark our system against prior photoelastic and vision-based tactile sensors, showing that our approach achieves competitive or superior results. section 6 explores key insights, including trade-offs in sensor construction and opportunities for improving both the models and the sensor. Finally, section 7 summarises the main contributions of this work. By pairing low-complexity hardware with task-specific deep learning, our system delivers accurate, real-time tactile sensing with strong potential for use in practical robotic applications.

2 State-of-the-Art

Early efforts to exploit the photoelastic effect for tactile sensing demonstrated its fundamental feasibility but were limited in scope and functionality. One of the earliest examples, by Bertholds et al. (1986) [5], used photoelasticity for pressure sensing via a long optical fiber combined with polarisers and photodetectors to measure stress-induced light changes. Dubey et al. (2006) [6] extended this idea with several photoelastic tactile sensors, including a dynamic system that detected both normal forces and incipient slip using a polarised light source and a photodiode. Their setup achieved a normal force range of up to 6 N and could detect slip speeds around 0.1 mm/s. While this work proved the concept of real-time slip sensing, it was limited to single-point measurements and could not resolve complex contact distributions.

Subsequent developments shifted toward camera-based imaging of fringe patterns to enable richer data capture. For example, Dubey et al. (2007) [7] demonstrated that a neural network could be trained on fringe images to estimate applied loads. Their system successfully mapped fringe orders to normal forces up to approximately 95 N with an error of around 2.8% (approximately 2.6 N absolute error) using a 1 MP camera. However, this work did not address tactile sensing directly but established that fringe order could predict force magnitude. Chung et al. (1998) [8] took a similar approach, employing a neural network with photoelastic imaging to estimate torque, achieving a low error of 0.4%. Despite their accuracy, these early systems focused on a single output (either force magnitude or torque), relied on relatively simple networks by today's standards, and typically assumed a known, centred contact position—failing to infer either contact location or object identity [3].

More recently, researchers have aimed to expand the scope of photoelastic tactile sensing to more practical, real-world applications. Hardware-based multi-axis sensing has become more common. For instance, Mitsuzuka et al. (2022) [4] developed a tactile sensor using a highly photoelastic polyurethane sheet paired with multiple LEDs and photodiodes. This setup enabled the measurement of both normal (Fz) and tangential (Ft) forces through stress-induced birefringence, with the force components separated via three distinct light paths. However, with only three photodiode readings, the system lacked the spatial resolution needed to detect contact location or infer object shape.

Camera-based designs have also matured. Mukashev et al. (2022) [3] introduced the PhotoElasticFinger—a robotic fingertip using soft silicone that produces photoelastic fringes. Their system estimated forces up to 8 N with RMSE of less than 2.5 N, confirming the potential of photoelastic imaging for robotic applications. Still, their sensor was limited to estimating the total force magnitude and did not provide contact location or shape recognition.

Other research has explored specialised use cases or materials. Takarada et al. (2021) [9] embedded photoelastic polyurethane into robotic fingertips to measure gripping force and classify object stiffness. Their system demonstrated how fringe patterns vary with object compliance—suitable for handling soft or delicate items. However, the method relied on average light intensity rather than full-frame image processing, limiting its spatial precision.

Liu et al. (2024) [10] introduced a fatigue-resistant mechanoresponsive hydrogel (FMCH) designed for tactile applications. Their sensor exhibited robust and consistent optical behavior across over 10,000 usage cycles and was capable of inferring object shape, location, stiffness, and pressure from reflected fringe patterns. While technically sophisticated, the system required custom hydrogel synthesis and a complex optical stack, including pre-stretching of the FMCH layer inside the sensor.

Complementing these works, Mitsuzuka et al. (2020) [11] presented a full camera-based photoelastic sensor using an engineered polymer film with enhanced birefringent properties. Their design incorporated a light source, polarisers, and a CMOS camera to visualise stress distributions. The system successfully handled both normal and shear loading, and could resolve contact positions and pressure directions—demonstrating practical feasibility for capturing distributed tactile signals.

Other implementations targeted different goals entirely. Nakamura et al. (2013) [12] used a layered photoelastic sensor to distinguish between human-perceivable surface textures, but the system did not attempt to measure contact forces. Likewise, applications beyond robotics—such as using fringe patterns for analyzing foot pressure in medical diagnostics [13] or building a pressure-sensitive touch interface on screens [14]—show the breadth of photoelastic sensing but do not address the combined inference of force, position, and shape.

Comparing capabilities. Previous photoelastic tactile sensors present trade-offs between simplicity and information richness. Systems using photodiodes [4] respond quickly but offer limited spatial detail. In contrast, camera-based systems offer full-field fringe imaging [3] that can in theory support richer inference, but historically required offline processing or predicted only a single tactile variable. To our knowledge, no prior sensor combines force estimation, contact localisation, and shape recognition in a unified framework, nor applies modern deep learning to decode photoelastic signals in real time. Many designs are constrained by narrow force ranges (typically below 10 N [3]), complex fabrication techniques (e.g., embedded fibres or multi-layered films), or evaluation under narrow conditions. This

leaves a clear gap in the research: a camera-based photoelastic tactile sensor that is easy to manufacture, supports multi-task inference, and generalises across conditions. Our system addresses this gap by combining a simple sensor design with a lightweight multi-task CNN capable of real-time, spatially resolved tactile prediction.

3 Methods

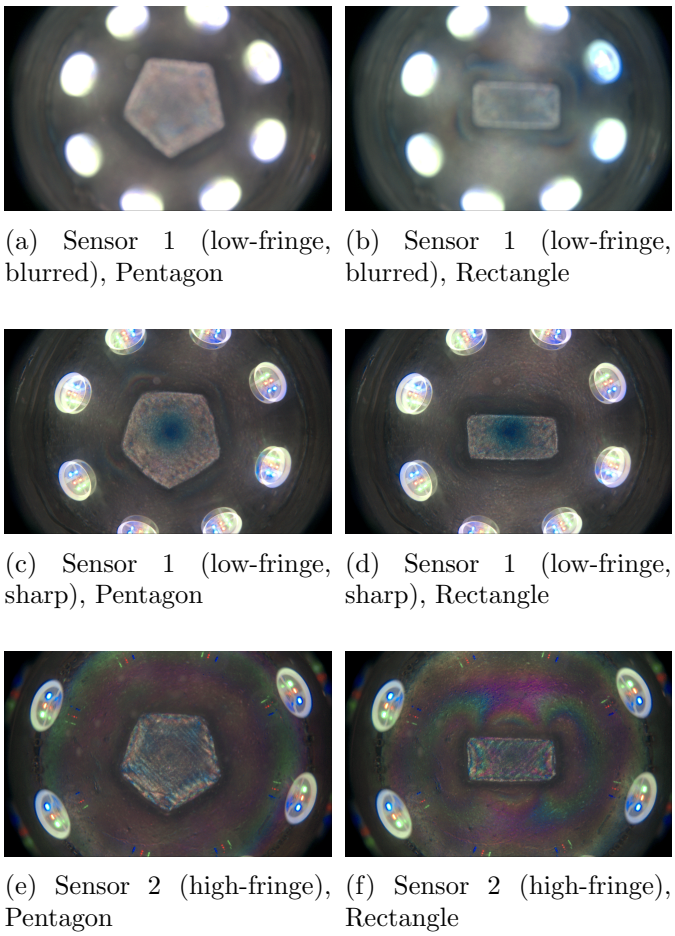
3.1 Sensor Design

The tactile sensor used in this project operates based on the principle of photoelasticity. When a transparent elastic material is subjected to an external force, internal stress causes the formation of visible fringe patterns—known as isochromatics. These patterns, illustrated in Figure 1, reveal how force is distributed across the surface and serve as a source of visual information. By analysing these fringe patterns, machine learning models can be trained to infer the underlying forces applied to the sensor. As shown in the same figure, the fringe patterns combined with the elasticity of the sensor also encode the shape of the indentor and the contact location. Further details on how the model learns to extract and predict these features are discussed in subsection 3.4.

Over the course of this project, we developed and evaluated seven distinct sensor prototypes, starting with designs that featured a pronounced curvature. These early sensors demonstrated strong potential for precise small-force measurements due to their stress-focusing geometry. However, while effective for force estimation, their geometry limited the range of additional tactile information we could extract. To achieve a more versatile sensing system—capable not only of force prediction but also of detecting contact location and identifying object shape—we shifted toward flatter designs. After experimenting with several curvatures, we ultimately selected a flat sensor configuration. This approach retained accurate force sensing while unlocking richer spatial features such as contact point localisation and shape determination, similar to the capabilities shown in GelSight systems [15].

Sensor architectures

We fabricated two photoelastic tactile sensors—**Sensor 1 (low fringe)** and **Sensor 2 (high fringe)**—whose sensing modules are labelled "1" in Figure 2. Both versions use a birefringent



(a) Sensor 1 (low-fringe, blurred), Pentagon
(b) Sensor 1 (low-fringe, blurred), Rectangle
(c) Sensor 1 (low-fringe, sharp), Pentagon
(d) Sensor 1 (low-fringe, sharp), Rectangle
(e) Sensor 2 (high-fringe), Pentagon
(f) Sensor 2 (high-fringe), Rectangle

Figure 1: Sample raw images from the three sensor/image quality combinations. Sensor 1 was recorded both with blurred (a,b) and sharp (c,d) focus settings. Sensor 2 consistently produces distinct photoelastic fringes (e,f). Each image shows an indentation in the center position (0,0) using either a pentagon or rectangle indentor.

polythioether (PTE) layer that reveals stress through visible photoelastic fringes (shown in Figure 1).

In conventional setups, the photoelastic material must be placed between two circular polarisers. However, since both the illumination and the camera are positioned on the same side, a reflective 200 µm adhesive layer mixed with silver pigment is applied to the object-facing side, as shown in Figure 2. A circular polariser is then placed on the camera side to complete the optical path and make the fringes visible.

The two sensors differ in how the photoelastic and supporting layers are structured:

- **Sensor 1 (low fringe)** is a laminated stack: a thin (1 mm) PTE film is bonded to a 2 cm soft

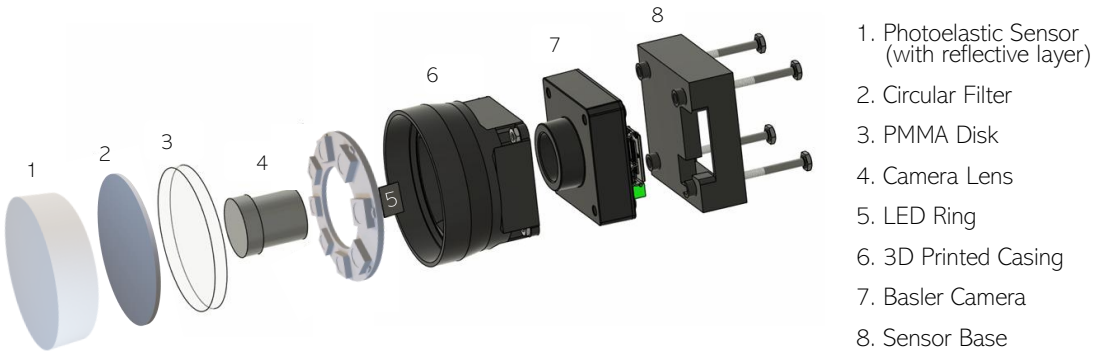


Figure 2: Sensor exploded view [16]

polyurethane layer for shape capture.

- **Sensor 2 (high fringe)** simplifies this by combining both into a single 2 cm soft photoelastic layer with enhanced fringe visibility.

Both variants share the same reflective coating, polariser film, and 5 mm PMMA base. The full preparation procedure and material details are described in Appendix A.2.

3.2 Measurement Setup and Data Acquisition

To collect training data, we developed a custom experimental setup capable of recording both the visual stress patterns and the corresponding ground-truth force information applied to the photoelastic sensor. The setup, shown in Figure 3, included a Basler daA1920-160uc RGB camera (S-Mount) [17] to collect the images mounted below the transparent sensor discussed in section 3.1, and a 6-axis ATI Mini40-E force/torque sensor [18], attached to the L bracket, connecting the stepper with the indenter. The camera captured fringe patterns from beneath the sensor, while the force sensor recorded applied forces and torques as the indentor pressed into the soft layer.

Both sensors contain a circular filter, and a reflective coating. This configuration allowed light to pass through the sensor, interact with internal stress fields, and produce visible photoelastic fringes that the camera can record. These fringes form the basis for visual force estimation and are explained in detail in subsection 3.1.

To control the indentation process, we used two orthogonally mounted Thorlabs stepper motors [19] to position the indentor in the x and z directions. The z-axis motor applied force, while the x-axis motor shifted the contact point laterally. A Thorlabs manual translation stage [20] was used for positioning along the y-axis. This configuration enabled precise 2D control over the location of indentation.

During data acquisition, 2 indentors shapes (pentagon and rectangle) were pressed into the sensor at different positions and force levels. The force/torque sensor recorded all six components of the applied load at 5 kHz, while the Basler camera captured RGB images of the resulting fringe patterns at 50 Hz. Each trial consisted of a time window during which the system continuously recorded synchronised streams of image and force data. This produced a sequence of frames, each with an image, a force vector, and metadata including the indentor shape and the (x, y) contact position relative to the sensor centre.

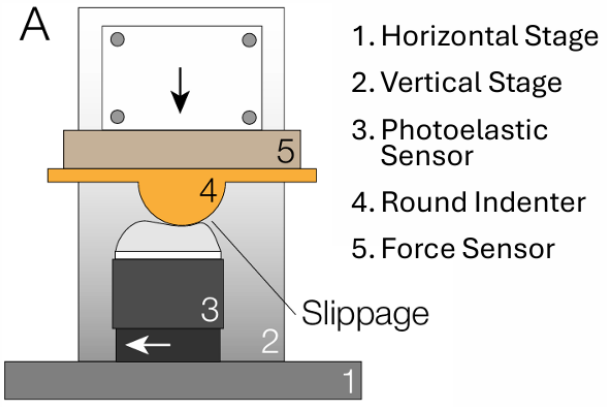


Figure 3: Measurement setup for recording force and fringe image data [16]

To account for noise and avoid labelling frames with negligible or unintentional contact, we applied a force threshold. Specifically, when the measured force magnitude dropped below 0.2 N, the sample was considered as having no meaningful indentation. In such cases, the indentor shape was set to `"None"`, and the contact coordinates were marked as `NaN`. This threshold was chosen based on the observation that the force sensor used for this data acqui-

sition exhibited force noise of up to approximately 0.2 N. Excluding these low-force readings helped reduce label noise and ensured that the model was only trained on valid contact events.

3.3 Dataset Construction and Augmentation

Why Augmentation Matters

To build a model that performs reliably under real-world conditions, it is essential to train it on data that reflects the variability and imperfections found in physical systems. In photoelastic tactile sensing, fringe patterns are sensitive to lighting, object indentation, etc. Augmentation allows us to simulate this variability without needing to physically collect new data for every condition. Throughout this project, we explored three different augmentation strategies, each chosen to test a specific trade-off between performance, realism, and computational cost.

Overview of the Three Augmentation Pipelines

We applied three distinct augmentation pipelines, each serving a different purpose in our study. A short summary is shown in 1 for clarity.

Table 1: Augmentation strategies used during training.

Augmentation Strategy	Description and Trade-offs
1. Grayscale with Noise	Converts the RGB image to grayscale and applies 10 augmentations per sample, including jitter and Gaussian noise. High augmentation diversity improves generalisation and robustness in real-world settings, but requires more computation.
2. RGB with Noise	Keeps the full colour image and adds the same noise types, but only 3 augmentations per sample. Less augmentation reduces robustness, though training is faster and less computationally intensive. Useful when resources are limited.
3. RGB without Noise	Uses clean RGB images with only 3 augmentations and no added noise. Represents an idealised scenario with fast training and clean data. However, poor generalisation makes it less reliable for deployment in variable environments.

To illustrate the core concept behind our augmentation process, we visualise the transformation sequence for the grayscale with noise pipeline in Figure 4. While this specific example reflects one of the three pipelines, the shared geometric transformations are common to all. Colour-specific and noise-related differences are discussed later.

Motivation Behind Each Pipeline

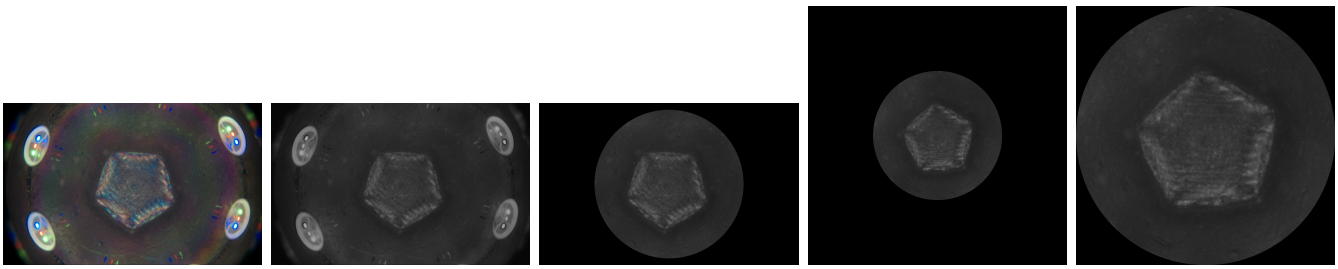
Each augmentation method was designed with a specific goal in mind. The grayscale pipeline was the first to be implemented and served as an initial strategy to test whether tactile information could be accurately extracted from fringe patterns without relying on color. Since the core photoelastic effect is based on light intensity variations and fringe structures—which are still visible in grayscale—the hypothesis was that colour might not be necessary for accurate force and shape prediction. Converting the image to grayscale also reduces the input complexity, enabling faster training and simplifying model debugging during early experimentation.

Additionally, we applied 10 augmentations per sample in the grayscale setup to introduce variation in lighting, noise, and orientation. This was intended to improve the model's robustness and generalization, especially in the early stages when the dataset was relatively small.

The two RGB pipelines were introduced to assess whether retaining colour information could enhance model performance, particularly for tasks that rely on subtle variations in fringe patterns which may be diminished in grayscale. One pipeline included the same noise and jitter augmentations used in the grayscale setup, simulating real-world variability. The other pipeline used clean, unaltered RGB images to evaluate performance under ideal, noise-free conditions.

Common Augmentation Steps

All pipelines share a number of preprocessing steps that ensure the augmented data remains physically accurate and consistent across the dataset. While the specific parameters and intensity of augmentation may vary by pipeline, the general process is similar. Figure 4 illustrates one representative example of this pipeline, demonstrating the core steps applied to a single input image. These core steps are as follows:



(a) Original (b) Grayscale+Noise (c) Shift+Mask (d) Rotation (45°) (e) Cropped

Figure 4: Visual walkthrough of the image augmentation process used in the grayscale with noise pipeline. Starting from a 1920×1200 RGB frame, the image is converted to grayscale and expanded to three channels. Brightness and contrast are then randomly jittered, followed by the addition of Gaussian noise (= 5). The image is horizontally shifted by 30 pixels to correct for optical misalignment, masked using a centered circular aperture with a diameter of 1100 pixels, randomly rotated between 0° and 360° around the true sensor center, and finally center-cropped to 1100×1100 pixels. While this example reflects the grayscale+noise pipeline, the same geometric transformations are applied consistently across all augmentation methods to preserve label validity for contact point, force, and shape.

1. **Center cropping:** The sensor is placed at the true center of the image to ensure that all rotations happen around the optical center.
2. **Rotation:** Each image is rotated randomly between 0° and 360°, with the contact point label adjusted accordingly to make the model as generalizable as possible.
3. **Horizontal shifting:** A fixed 30-pixel shift corrects for slight alignment errors between the camera and the sensor.
4. **Masking:** A circular mask is applied to remove irrelevant parts of the image outside the sensor area.
5. **Final cropping:** The resulting image is cropped to a fixed size for model input—980 pixels for Sensor 1 and 1100 pixels for Sensor 2.

These transformations ensure the labels (force, contact location, and shape) remain valid and consistent, while increasing the variety of conditions the model sees during training.

For models trained on augmented data with noise and jitter, these visual distortions are applied first—before any geometric transformations described above. Specifically, colour jitter randomly adjusts brightness and contrast to reflect variations that often occur in real-world imaging setups. In our sensor, these fluctuations can be caused by inconsistencies in the LED ring light—such as slight voltage variations or thermal drift—or by ambient light leaking from the outside world through the partially transparent elastomer of the sensor. Gaussian noise is added as random pixel-wise fluctuations, simulating image degradation introduced by the camera itself. This includes effects like low-light noise, thermal noise, and electronic interference. In the grayscale pipeline, these perturbations are applied after converting the image from RGB to grayscale, ensuring the resulting image maintains physical realism while still capturing noise characteristics. Together, these augmentations encourage the model to focus on stable fringe features rather than overfitting to ideal conditions, helping it generalise to a broader range of deployment scenarios.

Grayscale Image Processing

In the grayscale augmentation pipeline, each sensor image was converted to a single-channel intensity representation using `.convert("L")`. This operation computes the luminance of each pixel based on a weighted sum of the red, green, and blue colour components as shown in the following equation:

$$\text{Grayscale Intensity} = 0.299 \cdot \text{Red} + 0.587 \cdot \text{Green} + 0.114 \cdot \text{Blue} \quad (1)$$

This formula reflects the human eye's greater sensitivity to green and red light, giving more weight to those channels. After the image is converted to grayscale, we expand the result back into a 3-channel tensor using `repeat(3,1,1)`. This duplication preserves compatibility with CNNs expecting RGB inputs, such as the ResNet-18, without changing the model architecture.

While this approach simplifies the input by removing colour variation, it may also eliminate potentially useful features. In photoelasticity, colour fringes arise from stress-induced birefringence, and subtle chromatic differences can encode additional information about the magnitude and distribution of stress. Therefore, although grayscale conversion captures core structural information, it potentially does so at the cost of discarding meaningful optical cues.

Working with RGB Images

The RGB pipelines preserve all colour information in the image, which can be valuable for recognising complex patterns in fringe responses. The noisy version applies the same augmentation steps as the grayscale pipeline, but uses fewer copies (three instead of ten) to keep training time manageable. The noise-free RGB version skips jitter and Gaussian noise altogether. It was designed not for deployment, but to explore the best possible performance when the input is clean—such as in simulations or tightly controlled lab settings.

Sensor-Specific Cropping Sizes

Because the two sensors have different construction characteristics, we used different final cropping sizes. Sensor 2, which had a thicker photoelastic layer, shows more spread out light reflections near the edge, allowing us to use a wider 1100-pixel crop. Sensor 1, being more sensitive to reflections, was cropped to 980 pixels to exclude noisy borders and focus on the usable area.

Summary and Implications

These three augmentation pipelines were developed to explore trade-offs between accuracy, realism, and computational efficiency. The grayscale pipeline—with heavier augmentation—tests how well models perform under broader visual variation. The noisy RGB pipeline investigates whether retaining colour improves performance under similar augmentation conditions. The clean RGB version, in contrast, aims to reveal the best-case scenario when no artificial noise is added. Together, these strategies offer insight into how different data preprocessing choices affect tactile model performance. As tactile systems are scaled and deployed in real-world settings, understanding the implications of augmentation depth, colour retention, and noise simulation becomes crucial for designing models that are both robust and practical.

3.4 Model Architecture and Training

To learn from photoelastic fringe images, we implemented and compared two neural network architectures: a ResNet18-based model and a lightweight custom CNN. Each serves a distinct purpose.

ResNet-18 is a widely used baseline in computer vision, known for its residual connections that stabilize training in deeper networks [21].

In contrast, the **custom CNN** was specifically designed for low-latency inference in resource-constrained settings. With a significantly smaller parameter count than ResNet-18, it enables faster execution and reduced computational overhead. In live testing, the custom model consistently outperformed ResNet-18 in speed, achieving an average inference rate of 10.5 Hz compared to 9 Hz for ResNet-18. These benchmarks were obtained using only the CPU of an HP ZBook laptop, suggesting that substantial speed gains are possible with GPU acceleration.
Using both architectures allows us to explore the trade-off between general-purpose deep networks and task-specific lightweight designs. This comparison directly supports our investigation into how sensor design and image quality affect learnability and performance.

Both models were structured to solve a three-headed multitask learning problem. Given an input isochromatic fringe image, the network simultaneously predicts:

1. A 7-D force vector $[F_x, F_y, F_z, F_t, M_x, M_y, M_z]$,
2. The indenter shape class, and
3. The contact point coordinates (x, y).

Although the network predicts the full 7-dimensional force vector, the current loss function penalizes only the normal force component, F_z. This design choice enables straightforward future extensions: once ground truth data for torque and shear components become available, the loss can be adjusted accordingly without modifying the model architecture or retraining from scratch. For this project, the focus on F_z served both to reduce the number of required labels and to demonstrate that accurate normal force estimation is feasible using our sensor setup.

ResNet18-based model. This architecture adapts the standard ResNet18 by removing its final classification layer and appending three parallel output heads: a 7-unit regression head for the force vector, a softmax classification head for the indentor shape, and a 2-unit regression head for the contact point. The model is initialized from scratch (no pretraining) to ensure adaptation to the unique structure of photoelastic images.

Lightweight custom CNN. The custom model consists of four convolutional blocks, each with ReLU activations and max pooling. The number of channels doubles after each block (from 16 to 128), while the spatial resolution is halved. An adaptive average pooling layer compresses the feature map to $(128 \times 1 \times 1)$, which is then flattened and passed to the same three output heads used in the ResNet variant.

Training setup. Both models were trained for 40 epochs using the Adam optimizer with a learning rate of 1×10^{-3} and a batch size of 32. Images were downsampled and normalised. Depending on the chosen augmentation strategy, they were either converted to grayscale with additional noise or kept in full RGB with or without noise. The dataset was split into training (80%), validation (10%), and test (10%) sets using a fixed random seed to ensure consistency. Training was carried out on the DelftBlue supercomputer [22], using PyTorch's `DataParallel` API to support multiple GPUs.

Loss function. A composite multi-task loss function was used to train all three prediction heads simultaneously. This loss combines three components:

- Mean Squared Error (MSE) for the F_z force component
- Cross-Entropy Loss for the indentor shape classification
- MSE for the 2D contact point (x, y), applied only when both values are valid

These components are summed with equal weights. The total loss L for a single batch is computed as:

$$L = \mathrm{MSE}\big(F_z^{\mathrm{pred}}, F_z^{\mathrm{true}}\big) + \mathrm{CE}\big(\mathrm{Shape}^{\mathrm{pred}}, \mathrm{Shape}^{\mathrm{true}}\big) + \mathrm{MSE}_{\mathrm{valid}}\big((x,y)^{\mathrm{pred}}, (x,y)^{\mathrm{true}}\big) \quad (2)$$

The third loss term corresponds to the contact point prediction and is only computed for samples with valid (x, y) annotations. As explained in subsection 3.2, when the measured normal force F_z falls below 0.2 N, the signal is treated as noise and the sample is labeled with `NaN` contact coordinates and an indenter shape of `None`. To prevent the model from learning from unreliable labels, a binary mask is applied during loss computation to exclude these invalid contact points from the regression loss. Importantly, such samples are not discarded entirely. Even when the contact point is missing, the corresponding images are still used to compute the force regression and shape classification losses. This approach ensures that the model utilizes all available data while avoiding the introduction of label noise in tasks where ground truth is not meaningful.

To match the resolution of the physical sensors and reduce label noise, target values F_z and (x, y) were rounded to one decimal place. This rounding reflects the measurement accuracy of the hardware and helps stabilize training by avoiding gradients caused by insignificant fluctuations.

Model checkpoints were saved every 2 epochs, and both training and validation losses were tracked throughout. A final evaluation was performed on a test set using the same three metrics, reported as averages over all test samples. The detailed results are presented in section 4.

4 Results

This section presents a detailed evaluation of our tactile sensing system across multiple dimensions. We begin with a quantitative analysis of model performance, comparing different sensor designs, image sharpness levels, model architectures, and augmentation pipelines using standard metrics such as force error, contact localisation accuracy, and shape classification scores. We then introduce a complementary qualitative validation that explores how well the trained models generalise to real-world conditions using a live inference interface. Together, these evaluations provide a comprehensive view of the system's robustness, predictive accuracy, and practical viability.

4.1 Evaluation Metrics

To evaluate our tactile sensor and learning setup, we use a set of metrics that align with its three main prediction tasks: estimating normal force (F_z), localising the contact point, and classifying the indentor shape.

Force prediction is assessed using the Mean Squared Error (MSE) between the predicted and actual force values. In practice, we focus on the normal force component F_z. The MSE is calculated as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(F_{z,i}^{\text{pred}} - F_{z,i}^{\text{true}}i\right)^2 \tag{3}$$

which gives the average squared difference across all test samples. We use MSE because it emphasises larger errors more strongly, giving a clearer picture of the model's accuracy in predicting actual force values. A lower MSE directly implies a lower root-mean-square error, which in turn reflects how close the predictions are to the real force in Newtons. Since this metric is standard in regression, it also allows for easy comparisons with previous work that reports similar force-related metrics [3]. In our case, a low MSE on F_z means the sensor can consistently estimate the normal force with minimal deviation—an essential property for safe and stable robotic interaction.

Contact point localization is evaluated using the Mean Squared Error over the 2D contact coordinates (x, y) on the sensor surface. The formula follows:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^{\text{pred}} - x_i^{\text{true}}\right)^2 + \left(y_i^{\text{pred}} - y_i^{\text{true}}\right)^2 \tag{4}$$

For easier interpretation, this often gets converted into Root Mean Square Error (RMSE) measured in millimetres. This tells us how precisely the model can locate the point of contact. We favour MSE or RMSE here because it reflects the Euclidean distance between predicted and actual positions and provides a direct spatial error in physical units. A small contact point RMSE—typically within a millimeter—means the sensor can accurately locate where contact occurred. This is particularly useful for fine manipulation tasks or shape-based interactions. We selected this metric because it is intuitive, grounded in physical space, and comparable with values reported by other tactile sensors.

Indentor shape classification is evaluated using commonly used classification metrics: Accuracy, Precision, and F1-score. Accuracy gives the percentage of correctly predicted shapes out of all predictions, serving as an overall performance measure. However, when class distributions are uneven, accuracy on its own can be misleading. That is why we also report Precision—how many predicted instances of a given shape are correct—and F1-score, which balances Precision and Recall through their harmonic mean. The F1-score provides a single value that captures both types of classification error. These additional metrics help ensure the model performs well across all shape classes and doesn't just default to the most common one. For example, a model might reach high accuracy by always predicting the most frequent class, but that would show low Precision and F1 on less common shapes. Including these metrics helps us assess whether the model makes consistent, balanced predictions. Although shape classification is not usually addressed in previous photoelastic sensor work, these metrics allow us to benchmark this added functionality within our system.

4.2 Quantitative Results

Table 2 summarises the performance of both the ResNet-18 and the custom lightweight CNN across different sensor designs, image qualities, and augmentation pipelines. Each model was evaluated on its ability to predict force, contact location, and indenter shape. Beyond architecture comparisons, we assessed the effect of sensor structure—either a high-contrast integrated design (Sensor 2) producing more pronounced fringes or a layered low-fringe design (Sensor 1)—as well as image sharpness and augmentation strategies.

Impact of Image Quality: Blurred vs. Sharp
A clear performance gap emerges when comparing Sensor 1 results using blurred versus sharp images. Proper focus and optical clarity significantly enhance model performance across all tasks. For instance, the custom CNN* (see Table 2) sees its force RMSE drop from 1.5827 N to 0.5536 N, contact RMSE reduce from 3.5334 mm to 2.1048 mm, and F1-score improve from 0.6957 to 0.8739. ResNet-18 (see Table 2) shows similar gains, with improvements across force, localisation, and shape metrics. These results highlight the sensitivity of photoelastic fringe interpretation to image sharpness. Even in resource-constrained models, well-focused images enable richer feature extraction and markedly better performance.

Sensor Design: Layered vs. Integrated
Sensor 2 was designed to produce clearer and more distinct fringe patterns through an integrated optical stack. Nevertheless, experimental results show that Sensor 1—despite its simpler, layered construction and lower fringe contrast—achieves highly comparable performance. For example, ResNet-18 achieves its best results on Sensor 1 (sharp), with a force

Table 2: Model performance across sensor variants using ResNet-18 and Custom CNN. Metrics include force estimation (MSE, RMSE), contact localisation RMSE, and shape classification accuracy, precision, and F1-score.

Sensor Setup	Model	Force MSE (N^2)	Force RMSE (N)	Contact RMSE (mm)	Shape Acc.	Precision	F1-score
Sensor 1 (low-fringe, blurred)	ResNet-18	0.0521	0.2282	0.7595	0.9584	0.9578	0.9581
	ResNet-18*	0.5928	0.7699	1.0038	0.7884	0.8025	0.7933
	Custom CNN	1.4056	1.1855	0.9841	0.9167	0.9167	0.9033
	Custom CNN*	2.5051	1.5827	3.5334	0.7100	0.7305	0.6957
Sensor 1 (low-fringe, sharp)	ResNet-18	0.0213	0.1459	0.4462	0.9624	0.9611	0.9613
	ResNet-18*	0.0481	0.2193	0.8036	0.9500	0.9491	0.9484
	ResNet-18†	0.0488	0.2210	0.6857	0.9631	0.9672	0.9619
	Custom CNN	0.2428	0.4927	1.4683	0.9532	0.9576	0.9471
	Custom CNN*	0.3065	0.5536	2.1048	0.9013	0.8952	0.8739
	Custom CNN†	0.1854	0.4306	2.3124	0.9500	0.9494	0.9445
Sensor 2 (high-fringe, sharp)	ResNet-18	0.0298	0.1726	0.7528	0.9643	0.9685	0.9634
	ResNet-18*	0.0418	0.2044	1.0797	0.9577	0.9591	0.9564
	Custom CNN	0.3275	0.5722	2.3503	0.9274	0.9216	0.9152
	Custom CNN*	0.3740	0.6115	1.0797	0.9577	0.9591	0.9564

Note: Models marked with * or † were trained on 3 RGB-augmented images per sample (no grayscale). Models marked with † were additionally trained without added jitter or Gaussian noise and serve as noise-free benchmarks under ideal training conditions. All other models were trained with 10 grayscale-augmented images per sample and included noise-based regularisation.

RMSE of 0.1459 N compared to 0.1726 N on Sensor 2. The custom CNN follows a similar trend, reaching 0.4927 N on Sensor 1 versus 0.5722 N on Sensor 2. Although these differences are measurable, they remain small—particularly when rounded to one decimal place—rendering the two sensors nearly indistinguishable in practice. Given that Sensor 1 requires fewer fabrication steps and is easier to assemble, it raises an important question: is the marginal gain in accuracy from Sensor 2 worth the added complexity? In many practical settings, Sensor 2 may offer a more attractive trade-off between performance, cost, and manufacturability.

Augmentation Regime: Grayscale vs. RGB Across all setups, models trained with 10 grayscale augmentations generally outperform their RGB-trained counterparts with only 3 augmentations. However, the performance gap is not prohibitive. For instance, the custom CNN on Sensor 1 (sharp) (see Table 2) achieves an F1-score of 0.9471 in grayscale versus 0.8739 with RGB, and force RMSE rises modestly from 0.4927 N to 0.5536 N. That said, the RGB pipeline's efficiency is hard to ignore. Training the RGB-augmented models on a standard HP ZBook laptop takes less than half the time of the grayscale setup, which requires the same amount of time where the preprocessing takes place on a 48-core server and training on 4 GPUs.

To better understand the theoretical ceiling of RGB-based models, we introduce the † variants—models for Sensor 1 (low fringe, sharp), trained on RGB inputs without any added jitter or Gaussian noise. These configurations simulate an idealised environment. Under these controlled circumstances, the ResNet-18† (see Table 2) achieves a force RMSE of 0.2210 N and an F1-score of 0.9619, nearly matching the performance of the grayscale baseline (0.1459 N RMSE, 0.9613 F1). Similarly, the custom CNN† (see Table 2) reaches 0.4306 N RMSE and an F1-score of 0.9445— outperforming its noisy RGB counterpart (*).

While these results highlight the best-case performance for RGB models on this sensor, they should not be interpreted as representative of real-world reliability. In practical applications, sensor input is often affected by noise, such as lighting inconsistencies, all factors intentionally excluded from the models. Still, the findings are valuable: they highlight the upper performance limits of RGB models under tightly controlled or simulated conditions, such as robotic lab setups, enabling researchers to assess model headroom and refine architectures accordingly.

As datasets grow and tactile inference expands to include material recognition, torque estimation, or dynamic interactions, RGB pipelines remain a valuable option. Their low computational cost and quick turnaround times make them well-suited for large-scale training and field deployment, especially when paired with effective regularisation and domain-

specific fine-tuning.

Key Takeaways
Several important insights emerge from this analysis:

- **Sharp images dramatically improve performance**. Especially in layered sensors like Sensor 1, optical clarity has a measurable effect on all model outputs.
- **Sensor 1 performs on par with Sensor 2**. Despite having less fringe contrast, Sensor 1 matches or exceeds Sensor 2 in performance.
- **ResNet-18 is the most accurate model overall**. It consistently achieves the best force, contact, and shape predictions, though the custom CNN remains highly competitive—especially in lightweight deployments.
- **Grayscale augmentation yields peak performance, while RGB offers faster and more scalable training.** For real-time applications or large-scale datasets, RGB pipelines—when properly regularised—provide an attractive trade-off between training speed, computational cost, and predictive accuracy. Future work could explore the full potential of RGB inputs by applying the same 10-fold augmentation strategy used in the grayscale pipeline.
- **Noise-free RGB benchmarks validate model potential**. The † results demonstrate that under ideal conditions, RGB models can achieve high accuracy with minimal augmentation and no domain-specific preprocessing.

Altogether, the results illustrate how thoughtful sensor design, image clarity, and augmentation choices can unlock high-quality tactile prediction—even with compact models. These findings lay the groundwork for cost-effective, deployable tactile systems that do not compromise on performance.

4.3 Qualitative Evaluation: Live Inference Validation

To assess how the trained models perform in a real-world setting, we conducted an additional qualitative validation using the live inference interface developed for this project. This test was not part of the controlled dataset evaluation but offers valuable insight into real-time prediction accuracy under natural conditions.

The interface displays both the raw camera feed and the preprocessed input fed into the neural network—cropped, shifted, and masked identically to the training augmentations, but without added noise or jitter. Predictions for normal force (F_z), contact location (x, y), and shape class are then visualised through a live graphical interface. Using this tool, we manually tested all trained models by pressing different shapes onto the sensor surface, without using a fixed setup or alignment stage. Instead, indenter placement varied in orientation and position, emulating realistic use.

We tested whether the shape classification matched the true shape and whether the predicted contact location was consistent with the physical point of contact. When indenters were pressed lower on the sensor, the y-coordinate of the prediction was expected to be negative, as the middle of the sensor is the (0,0) point. Similarly, upper presses should yield a positive y-value. The same validation logic was applied for the x-coordinate. We further verified that contacts near the centre produced values close to $(0, 0)$, while those near the edge had magnitudes exceeding 10 mm.

These tests were repeated under two lighting conditions: in a dark room with minimal ambient light, and in a bright environment with directed lighting on the sensor. Across all scenarios, the ResNet-18 models consistently performed well. They correctly identified the shape, localised the contact position with accurate sign and magnitude, and handled the "None" class (no contact) robustly. In contrast, the custom CNN models sometimes struggled to distinguish between similar shapes and showed more variability in contact location predictions. Notably, the custom CNN often defaulted to one dominant shape class, misclassifying both pentagon and rectangle shapes as the same label when the predicted force exceeded 3 N, which was determined by the model. At lower predicted forces (below 3 N), the model performed more accurately in distinguishing shapes.

We attempted to validate force prediction as well by placing known weights on the indenters. However, this proved unreliable, as the models were trained exclusively on data collected with vertical indentations, assuming a parallel orientation between the indenter and the sensor. Without exposure to angular indentations during training, generalisation to off-axis loading was limited. This suggests a

promising avenue for future improvement: training the model on a more diverse set of loading orientations, or incorporating all six force-torque components into the loss function, allowing it to train on more parameters, which may significantly enhance robustness under natural conditions.

In summary, this qualitative evaluation highlights the practical reliability of our ResNet-18 models for live tactile inference. While quantitative results remain the benchmark, live testing reveals the strength of our multi-task setup in real-world use, especially in shape and location prediction under varying lighting and contact conditions. These observations, while informal, reinforce the quantitative results: the ResNet-18 models are more robust across lighting and contact conditions, while the custom CNN shows more sensitivity to contact orientation and shape ambiguity.

5 Quantitative Comparison with State-of-the-Art

To situate our sensor within the broader landscape of tactile sensing, we compare its performance with systems introduced after 2005. This includes mostly photoelastic sensors, which share the same underlying optical principle, and also includes a vision-based design, namely GelSight. GelSight sensors are frequently referenced in the literature for their high spatial resolution and have served as a reference point in numerous tactile sensing studies [15]. Including it in our evaluation offers a clear benchmark beyond the scope of traditional photoelastic sensors.

Force Estimation

Force sensing is the most widely implemented capability in photoelastic designs. PhotoElasticFinger [3], for example, reliably detects normal forces ranging from 0.5–8 N, with errors between 0.91 and 2.5 N. Mitsuzuka et al. [4] extended this concept by measuring both normal and tangential forces through a specialised photoelastic sheet with embedded photodiodes. Their reported tangential force errors were under 4% of full scale. Our sensor operates across a broader 0–9 N range and achieves a force estimation error of approximately 1.6% full-scale, demonstrating competitive performance without requiring complex hardware.

Contact Localization

Few photoelastic systems report contact localisation. Most, such as the works of Dubey [6] and Takarada [9], treat the sensor as a single tactile unit. In contrast, our system estimates the location of contact on the sensing surface, achieving a root-mean-square error of 0.4462 mm.

Shape Recognition

Shape detection is virtually absent from current photoelastic sensors. PhotoElasticFinger, for example, reports total force but does not distinguish shape or position. GelSight excels in this domain, reconstructing surface topography at a spatial resolution of 20–30 μm. Although our approach does not recreate a full 3D profile, it classifies contact shape from fringe patterns with an estimated 96% accuracy. This makes our system the only one in the photoelastic category to offer shape recognition.

Table 3: Comparison of our best-performing photoelastic tactile sensor with state-of-the-art tactile sensors developed after 2005. Metrics include force estimation, contact localisation, and shape classification accuracy. Dashes (–) indicate unavailable or unreported data.
[1] Photoelastic sensors. [2] Vision-based tactile sensors (non-photoelastic).

Sensor*	**Force Range (N)**	**Force Error**	**Contact RMSE (mm)**	**Shape Accuracy (%) or Res.(μm)**
This Work[1] (Sensor 1, low fringe, sharp)	0–9	RMSE: 0.1459 N	0.4462	Shape determination accuracy: 96.24%
PhotoElasticFinger [3][1]	0.5–8	RMSE: 0.9–2.5 N	–	–
Mitsuzuka et al. (2022) [4][1]	4.1–9.6	$\leq$4% (tangential force)	–	–
Takarada et al. (2021) [9][1]	0.1–10	Resolution: 0.1	–	–
Mitsuzuka et al. (2020) [11][1]	0.02–10	$\sim$2% at 0.1 N	–	–
Dubey et al. (2006) [6][1]	0–6	–	–	–
GelSight [15][2]	0–20	RMSE: 0.668–1.856 N	–	spatial resolution: 20–30 μm

Summary and Context

Our sensor stands out for combining all three sensing tasks—normal force, contact position, and contact shape—within a single compact system. While prior work has often specialised in one modality, our approach is enabled by the use of a multitask deep neural network trained on fringe images. This allows us to extract multiple forms of tactile information without redesigning the hardware. Unlike GelSight, which requires post-processing for force estimation, our system infers it directly from the image. And unlike many earlier photoelastic systems, we are able to perform spatial sensing in real time.

Taken together, these capabilities show that our sensor bridges the gap between high-resolution tactile vision systems and the simplicity of photoelastic hardware. It offers a low-cost and scalable solution for multimodal tactile feedback—paving the way for its use in robotic manipulation, soft prosthetics, or embedded applications.

6 Discussion

6.1 Key Findings and Contributions

This work presents a modular, low-cost tactile sensing system based on the photoelastic effect, capable of simultaneously estimating contact force, location, and shape from single fringe images. Using synchronised high-frequency force data and structured augmentation, we trained a lightweight multi-task CNN and a ResNet-18 to make real-time predictions under different sensor configurations and image conditions. Notably, we observed that sharp images—whether from fringe-rich or fringe-sparse sensors—enabled the models to learn meaningful features. While grayscale images with extensive augmentation yielded the highest baseline accuracy, RGB pipelines offered a compelling trade-off between speed, flexibility, and resource usage. The results support the feasibility of photoelastic sensing for practical, real-time tactile feedback in robotics.

6.2 Fringe Quality and Image Clarity

Our results suggest that well-formed, high-contrast fringe patterns do improve model accuracy but are not strictly necessary for effective prediction. Sensor 2, with its integrated design and enhanced fringe visibility, was expected to offer a distinct performance advantage. However, sharp images from Sensor 1 performed comparably—often surpassing Sensor 2—despite its layered structure and lower fringe contrast. The clearest example is the ResNet-18 model on sharp Sensor 1 grayscale input, achieving a force RMSE of 0.1459 N and shape F1-score of 0.9613. In contrast, blurred images significantly degraded performance across all metrics, reinforcing the need for precise focusing of the lens. These findings imply that while fringe visibility contributes to performance, optical clarity is more critical. This insight allows for more flexible sensor fabrication, reducing reliance on high-fringe materials without sacrificing accuracy.

6.3 Design and Deployment Trade-Offs

Although Sensor 2 was designed with a unified optical stack to improve birefringence visibility, experimental results show that Sensor 1, despite its layered construction and lower fringe contrast, performs equally well—or in some cases better—when image sharpness is ensured. In particular, models trained on sharp grayscale images from Sensor 1 consistently achieved the lowest force prediction error (RMSE of 0.1459 N) and highest shape classification accuracy (F1-score of 0.9613). These findings indicate that high fringe visibility is not strictly necessary for accurate prediction, as long as the optical clarity is maintained.

From a deployment perspective, the two sensors offer distinct trade-offs. Sensor 2 simplifies the fabrication process by integrating the sensing and support layers into a single soft photoelastic gel, making it easier to produce at scale.

These results suggest that in real-world tactile systems, performance does not hinge solely on maximising fringe visibility. Instead, the interplay between optical clarity, sensor robustness, and manufacturing simplicity must be considered. The choice between Sensor 1 and Sensor 2 thus depends on application-specific priorities: Sensor 1 for maximum performance and robustness; Sensor 2 for ease of manufacturing and cost-efficiency.

6.4 Augmentation and Model Robustness

The choice of augmentation pipeline significantly influenced training efficiency and model generalization. The grayscale setup, which applied 10 diverse augmentations per image, performed best under realistic noise and lighting conditions. However, RGB pipelines—despite using only three augmentations—achieved surprisingly competitive results.

The noise-free RGB variant of Sensor 1 with sharp images achieved the lowest force MSE ($0.0488\,\mathrm{N}^2$) and highest shape classification accuracy (96.3%), providing an upper bound on model performance under ideal conditions. Importantly, this configuration does not reflect deployment settings, as it excludes jitter and Gaussian noise typically encountered in real-world images. Nevertheless, it serves as a valuable reference point, revealing the model's learning capacity in controlled environments.

On the other hand, the noisy RGB pipeline—which mirrors the grayscale augmentation logic but preserves colour—offered an attractive balance between speed and realism. Models trained with this method completed significantly faster and required fewer compute resources, making them practical for rapid experimentation. Altogether, our findings align with broader literature: augmentation strategies can regularise models, improve robustness, and reduce overfitting yet they must be matched to deployment constraints.

6.5 Advantages

This work introduces a tactile sensing system that stands out for its simplicity, versatility, and reliable performance. One of its key strengths is its low cost. Each sensor can be assembled for under €200 using standard, easily sourced materials. The modular casing design allows users to quickly replace the sensing material, making the setup ideal for rapid prototyping, repairs, and long-term use.

In terms of responsiveness, the sensor captures images at 50 Hz, and the lightweight multi-task model is capable of real-time inference. In practice, when deployed on a standard laptop without GPU acceleration, the full pipeline—including image capture, model prediction, and visualisation via a GUI—currently runs at approximately 10 Hz. While this already supports basic interaction and feedback, performance is expected to improve significantly with GPU-based inference or model optimisation for embedded hardware. Exploring techniques such as model pruning or quantisation could further increase speed while reducing memory load.

The sensor also demonstrates strong robustness. Despite differences in fringe clarity and internal design, both Sensor 1 and Sensor 2 produced reliable results, especially when the image quality was sharp. This indicates that the system tolerates moderate variation in manufacturing, lighting, or material properties, reducing the need for extensive recalibration across different sensor batches.

Efficiency in data collection and model training is achieved through a well-structured augmentation pipeline. By applying transformations such as rotation, noise injection, and brightness changes, we expanded the dataset significantly while preserving the original labels. This helped the model learn under diverse visual conditions without requiring additional time-consuming recordings.
Finally, the models themselves are compact and fast. While ResNet-18 delivered the highest overall accuracy, the custom lightweight CNN achieved similar results with far fewer parameters, making it suitable for deployment on resource-limited platforms such as mobile robots or wearable devices.
Altogether, these advantages make the system both practical and scalable. Its low cost, real-time performance, and adaptability to hardware constraints position it as a promising solution for tactile sensing in robotics, prosthetics, and interactive systems.

6.6 Limitations

While the system shows promise, several limitations remain. First, bright reflections near the sensor's edge reduce fringe visibility and can introduce noise in the model's predictions. This limits the effective sensing area. Second, the soft polyurethane surface will degrade over time due to scratches or repeated indentation, negatively affecting image quality. Third, the current dataset only includes normal forces; while the model architecture supports full 7D force prediction, including shear and torque components. Lastly, shape classification was limited to three classes: pentagon, rectangle, and "None" (no contact). Expanding this set is necessary for broader object recognition and generalisation. This paper presents the models that are able to determine forces from these fringes without the need of complex equations based on the material and creation of each sensor.

6.7 Future Work

Finally, generating data that **includes shear and rotational forces** would allow the model to fully leverage its multitask output head. Since the architecture already predicts the full 7-dimensional force vector, adding such labels would only require modifying the loss function, not the model itself. Beyond expanding the output space, this enhancement would also improve real-world applicability:

the model would no longer rely on the assumption that the object is pressed perpendicularly onto the sensor, enabling more accurate force predictions under varied orientations and more natural, unconstrained contact conditions.

As a more ambitious direction, future work could explore training the model on physics-based ground truth using **stress maps** derived from well-characterised materials. Although this would increase the complexity of the setup and calibration, it could yield more accurate predictions. This direction has been suggested as better suited for a longer-term project.

Expanding the model's capability to **reconstruct unfamiliar shapes** is another promising direction. Preliminary tests showed that the sensor can detect fine structural variations—such as cracks, dents, or surface textures—suggesting potential applications in inspection and quality control. For example, the system was able to visualise the stitching on denim fabric, indicating a spatial resolution comparable to GelSight. Importantly, by focusing on reconstructing the geometry or estimating the centre of contact from arbitrary shapes, the model could **infer optimal grasping points** without needing to explicitly classify each object. This would eliminate the need for extensive training on every possible shape or class. Instead of stating "this is a rectangle" or "this is a circle," the system could determine where to grasp based on physical interaction cues—supporting more generalised and adaptable robotic manipulation.

One promising direction is to **incorporate temporal information** from video sequences. Rather than treating each frame independently, the model could analyse how fringe patterns change over time. This temporal context would enable the detection of dynamic events such as incipient slip, object rolling, or gradual force buildup—phenomena that are difficult to infer from single images. By capturing and interpreting sequences of frames, the system could track how contact forces evolve and make more informed predictions about physical interactions.

Another important step is to more extensively **evaluate the system in real-world manipulation tasks**, such as robotic grasping, object placement, or tool use. These scenarios introduce more complex conditions—variable lighting, multi-object interactions, and uncontrolled contact orientations—that would test the model's ability to generalise beyond what is currently tested. Creating a dataset from these real-world tasks could further support domain-specific fine-tuning, allowing the system to adapt to specific applications while maintaining robust and accurate tactile inference.

Increasing the input resolution using square-format cameras (e.g., 1920×1920 pixels) would also allow a greater portion of the sensor area to be used for predictions. However, this could introduce new challenges, including increased noise due to reflections, edge artefacts, or material inconsistencies—especially near the periphery of the sensor.

A solution to this could be redesigning the illumination system to **distribute light more evenly** across the sensor, which could eliminate bright LED spots visible in some fringe images (see Figure 1), thereby expanding the usable area for inference.

Continued **model exploration** remains valuable. Testing different machine learning model architectures may lead to faster or more accurate alternatives, particularly for deployment on embedded systems. Collecting additional data—especially for underrepresented shapes or contact regions—could also strengthen generalisation.

Long-term durability is another aspect worth studying. **Quantifying** how **surface wear** affects sensing fidelity over time could help establish lifespan estimates and usage guidelines. For comparison, the GelSight sensor has been shown to endure up to 1000 repeated coin presses before replacement is required—a similar benchmark could guide future evaluation of these sensors' robustness [23].

7 Conclusion

This work introduces a practical and scalable approach to camera-based tactile sensing using photoelastic fringe patterns. By designing two sensor variants—one with a layered structure and one with an integrated high-fringe layout—and pairing them with a lightweight multi-task CNN, we demonstrate how visual stress responses can be turned into rich tactile information in real time. Our models accurately estimate normal force, contact position, and shape from a single image, making this one of the first tactile systems to achieve full-field multi-task inference using photoelasticity.

Through careful comparisons, we find that image sharpness plays a critical role in model performance. Interestingly, even Sensor 1, which produces fewer visible fringes, performs on par with the more distinguished fringes generated by Sensor 2 when well-focused. This suggests that simple sensors, when paired with clear imaging and good data preprocessing, can deliver high accuracy without the need for intricate fabrication. Our findings also emphasise that fringe clarity helps, but is not strictly necessary—blurred and low-contrast images still yield usable predictions, particularly when paired with robust augmentation.

We explored three distinct augmentation pipelines to study how colour, noise, and augmentation depth influence generalisation. While grayscale images with 10 augmentations produced the strongest results, RGB models trained with fewer augmentations showed surprisingly competitive performance. This trade-off between performance and efficiency is especially relevant for fast deployment on limited hardware. Furthermore, our RGB pipeline without added noise reveals the upper limit of what the model can learn under ideal conditions, though we stress that such setups are unrealistic for deployment and serve primarily as controlled benchmarks.

A key strength of our system lies in its accessibility. The full sensing stack—from sensor construction to model inference—can be built at low cost using off-the-shelf parts, with predictions running on a standard laptop between 9 and 10.5 Hz. While not yet real-time, this performance is likely to scale significantly with GPU acceleration or deployment on embedded platforms. The modular architecture and open pipeline design make it easy to adapt to new use cases, shapes, or sensor geometries.

Together, these results demonstrate that photoelastic fringe patterns offer a rich, underutilised signal for tactile sensing. With the right design, even simple hardware can yield high-resolution, low-latency tactile feedback. This opens up new possibilities for robot hands, prosthetics, and interactive systems that need fine-grained control over contact forces. Looking ahead, the system can be expanded to support torque, shear, or even contact material estimation, paving the way for real-world, multimodal tactile intelligence.

8 Acknowledgements

I thank Georgy Filonenko for his continuous feedback, invaluable learning moments, and support in connecting with the right people throughout the project. I am grateful to Agnes Kereszturi for designing and fabricating the sensors, and for thoughtfully implementing iterative feedback. I thank Giuseppe Vitrani and Michaël Wiertlewski for their weekly input, and Michaël for allowing me to use his Tactile Machines Lab, which enabled the development of the data acquisition setup. I also thank Gerwin Smit for his feedback on the project and the thesis. In addition, I thank Sid Kumar for his guidance on model development and for advising to postpone stress map integration to a longer-term effort. Lastly, I thank Simon Yoon for his feedback on the thesis.

References

[1] S. Dong, W. Yuan, and E. H. Adelson, "Improved gelsight tactile sensor for measuring geometry and slip," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 137–144, Sep. 1, 2017. DOI: 10.1109/iros.2017.8202149. [Online]. Available: https://arxiv.org/abs/1708.00922.

[2] Y. Li, W. Yuan, R. Kramer, and E. Adelson, "Sensing and recognizing surface textures using gelsight," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6609387.

[3] D. Mukashev *et al.*, "Photoelasticfinger: Robot tactile fingertip based on photoelastic effect," *Sensors*, vol. 22, no. 18, p. 6807, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/18/6807.

[4] M. Mitsuzuka, J. Takarada, I. Kawahara, *et al.*, "Application of High-Photoelasticity Polyurethane to Tactile Sensor for Robot Hands," *Polymers*, vol. 14, no. 23, p. 5057, Nov. 2022. DOI: 10.3390/polym14235057. [Online]. Available: https://doi.org/10.3390/polym14235057.

[5] A. Bertholds and R. Dändliker, "High-resolution photoelastic pressure sensor using low-birefringence fiber," *Applied Optics*, vol. 25, no. 3, p. 340, Feb. 1, 1986. DOI: 10.1364/ao.25.000340. [Online]. Available: https://doi.org/10.1364/ao.25.000340.

[6] V. N. Dubey and R. M. Crowder, "A dynamic tactile sensor on photoelastic effect," *Sensors and Actuators A Physical*, vol. 128, no. 2, pp. 217–224, Mar. 1, 2006. DOI: 10.1016/j.sna.2006.01.040. [Online]. Available: https://doi.org/10.1016/j.sna.2006.01.040.

[7] V. N. Dubey, G. S. Grewal, and D. J. Claremont, "Load extraction from photoelastic images using neural networks," *Experimental Mechanics*, vol. 47, no. 2, pp. 263–270, Jan. 5, 2007. DOI: 10.1007/s11340-006-9002-z. [Online]. Available: https://doi.org/10.1007/s11340-006-9002-z.

[8] D. Chung, F. L. Merat, F. M. Discenzo, and J. S. Harris, "Neural net based torque sensor using birefringent materials," *Sensors and Actuators A Physical*, vol. 70, no. 3, pp. 243–249, Oct. 1, 1998. DOI: 10.1016/s0924-4247(98)00147-2. [Online]. Available: https://doi.org/10.1016/s0924-4247(98)00147-2.

[9] J. Takarada, M. Mitsuzuka, Y. Sugino, *et al.*, "Evaluation of gripping sensor using polyurethane with high photoelastic constant," *Japanese Journal of Applied Physics*, vol. 60, no. SF, SFFD03, Aug. 27, 2021. DOI: 10.35848/1347-4065/ac2216. [Online]. Available: https://doi.org/10.35848/1347-4065/ac2216.

[10] J. Liu, W. Li, S. Yu, S. Blanchard, and S. Lin, "Fatigue-resistant mechanoresponsive color-changing hydrogels for vision-based tactile robots," *Advanced Materials*, Sep. 27, 2024. DOI: 10.1002/adma.202407925. [Online]. Available: https://doi.org/10.1002/adma.202407925.

[11] M. Mitsuzuka, Y. Kinbara, M. Fukuhara, *et al.*, "Relationship between photoelasticity of polyurethane and dielectric anisotropy of diisocyanate, and application of high-photoelasticity polyurethane to tactile sensor for robot hands," *Polymers*, vol. 13, no. 1, p. 143, Dec. 31, 2020. DOI: 10.3390/polym13010143. [Online]. Available: https://doi.org/10.3390/polym13010143.

[12] T. Nakamura, F. Kimura, and A. Yamamoto, "A photoelastic tactile sensor to measure contact pressure distributions on object surfaces," *Journal of Robotics and Mechatronics*, vol. 25, no. 2, pp. 355–363, Apr. 20, 2013. DOI: 10.20965/jrm.2013.p0355. [Online]. Available: https://doi.org/10.20965/jrm.2013.p0355.

[13] V. N. Dubey and G. S. Grewal, "Efficacy of photoelasticity in developing whole-field imaging sensors," *Optics and Lasers in Engineering*, vol. 48, no. 3, pp. 288–294, 2010, ISSN: 0143-8166. DOI: https://doi.org/10.1016/j.optlaseng.2009.11.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143816609002802.

[14] T. Sato, H. Mamiya, H. Koike, and K. Fukuchi, "Photoelastictouch: Transparent rubbery tangible interface using an lcd and photoelasticity," in *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '09, Victoria, BC, Canada: Association for Computing Machinery, 2009, pp. 43–50, ISBN: 9781605587455. DOI: 10.1145/1622176.1622185. [Online]. Available: https://doi.org/10.1145/1622176.1622185.

[15] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, 2017, ISSN: 1424-8220. DOI: 10.3390/s17122762. [Online]. Available: https://www.mdpi.com/1424-8220/17/12/2762.

[16] G. Vitrani, B. Pasquale2, and Wiertlewski, "GitHub - DoongLi/ICRA2025-Paper-List: ICRA2025 Paper List," *ShadowTac: dense measurement of shear and normal deformation of a tactile membrane from colored shadows*, [Online]. Available: https://github.com/DoongLi/ICRA2025-Paper-List.

[17] Basler AG. "Daa1920-160um (cs-mount) — basler ag." (), [Online]. Available: https://www.baslerweb.com/en/shop/daa1920-160um-cs-mount/.

[18] A. I. Automation and A. I. Automation. "Ati industrial automation: F/t sensor mini40." (), [Online]. Available: https://www.ati-ia.com/products/ft/ft_models.aspx?id=Mini40.

[19] *Thorlabs - NRT150/M 150 mm Motorized Linear Translation Stage, Stepper Motor, M6 Taps*. [Online]. Available: https://www.thorlabs.com/thorProduct.cfm?partNumber=NRT150/M.

[20] "1." (), [Online]. Available: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=706.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[22] D. H. P. C. C. (DHPC), *DelftBlue Supercomputer (Phase 2)*, https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2, 2024.

[23] GelSight, Inc., "Gelsight mini," 2024. [Online]. Available: https://www.gelsight.com/wp-content/uploads/productsheet/Mini/GS_Mini_4.3.24.pdf.

Appendices

A Sensor Assembly

This appendix gives a reproducible, step-by-step guide to rebuilding the photoelastic tactile sensor (exploded view in Figure 2).

1. **Bill of Materials** (§A.1)
2. **Layer-by-Layer Build** (§A.2)
3. **3-D-Printed Housing** (§A.3)

A.1 Bill of Materials

Table 4: Bill of materials for one sensor unit.

#	Component	Vendor / Part	Qty	Unit (€)
1	Basler daA1920–160*uc* RGB camera (S-mount)	Basler AG [6]	1	180
2	PMMA base, 60 mm Ø × 5 mm	CNC-machined	1	0.5
3	Polyurethane (Clear Flex 30) layer, 3cm	Smooth-On	1	0.5
4	Silver Mylar reflective film	-	1	0.1
5	White LED ring, 24 mm ID, 12 px	-	1	6
6	circular polarised filter	-	1	0.5
6	3-D-printed housing (PET-G)	—	1	—
7	M3 × 10 mm screws + inserts	ISO 7380	6	0.15

A.2 Layer-by-Layer Build

Both sensors were created using an open silicone mold for the preparation of the sensory units with a cylinder shaped cavity. The mold was made by casting silicone on a positive sample made of PMMA.

Version 1 — layered design: The different layers for this sensor are shown in Table 5. First, the photoelastic layer was made by pouring the mixture of acrylic and thiol monomers into the mould and letting them fully react and crosslink at room temperature. A transparent polymeric film was formed in this way in the bottom of the mould. As a second step, the commercial polyurethane formulation was mixed from the two components according to the instruction of the supplier. The mixture was degassed under vacuum to remove any air bubbles, poured in the mould on top of the photoelastic film and then let fully cure for a few days at room temperature. After this, the transparent polymeric sensor body was removed from the mould. To make a reflective layer, a commercial transparent elastic adhesive formulation was mixed with a mica-based silver pigment. A film with 200 micron thickness of the uncured fresh mixture was made by a hand film applicator on a smooth PE surface. Immediately after this the sensor body was placed on the film, the photoelastic layer facing and directly in contact with the reflective layer. The sensor body was cautiously pushed into the reflective layer to remove any trapped air between the two materials. The adhesive formulation was let to cure at room temperature for a day to form a silver colored thin film coating on the sensor. After this step the sensor was carefully removed from the PE base. Finally, the circular polariser film was attached to the polyurethane supporting layer by using a commercial transparent adhesive. On top of the filter a transparent PMMA hard layer was also fixed with the same adhesive.

Table 5: Layer stack of **Sensor 1** (object side → camera side).

#	Layer	Thickness	Material / description
1	Reflective layer	200 µm	Flexible adhesive mixed with silver pigment
2	Sensory layer	1 mm	Photoelastic polythio-ether film
3	Soft supporting layer	2 cm	Transparent polyurethane
4	Polariser	film	Circular polariser
5	Solid base	5 mm	PMMA plate

Version 2 — integrated design: In case of the improved second version of the sensor the thin photoelastic and the polyurethane supporting layers were merged (as shown in Table 6), therefore the preparation method was simplified. The new soft photoelastic gel material was casted in one single step in the silicone mould. The mixture of acrylic and thiol monomers was first degassed under vacuum to remove any air bubbles then poured into the silicone mould. The crosslinking reaction was accelerated by using elevated temperature on a hot plate at 80°C. After 3-4 hours the mixture was fully cured and the soft polymeric sensor was ready for further processing, as described before, adding the reflective layer, the circular polariser filter and the PMMA hard layer.

Table 6: Layer stack of **Sensor 2** (object side → camera side).

#	Layer	Thickness	Material / description
1	Reflective layer	200 µm	Flexible adhesive mixed with silver pigment
2	Soft sensory layer	2 cm	Photoelastic polythio-ether gel
3	Polariser	film	Circular polariser
4	Solid base	5 mm	PMMA plate

A.3 3-D-Printed Housing and parts

STL/STEP files and exact part specifications are archived at `https://github.com/yannickser/Real-Time-Multi-Task-Tactile-Sensing-Using-Photoelastic-Fringe-Patterns-and-Deep-Learning`. Figure 2 shows the CAD assembly.

B Code Repository

All source code, pre-trained models, and helper scripts needed to reproduce the experiments in this thesis are archived in the public repository

`https://github.com/yannickser/Real-Time-Multi-Task-Tactile-Sensing-Using-Photoelastic-Fringe-Patterns-and-Deep-Learning`

The root directory of the repository contains a comprehensive README.md file that documents the entire codebase, including data acquisition, data augmentation and preprocessing, model training, and live sensor testing.