Fisheye transformation enhances deep-learning-based single-cell phenotyping by including cellular microenvironment

Toth, Timea

2022-12

Fisheye transformation enhances deep-learning-based single-cell phenotyping by including cellular microenvironment

Graphical abstract
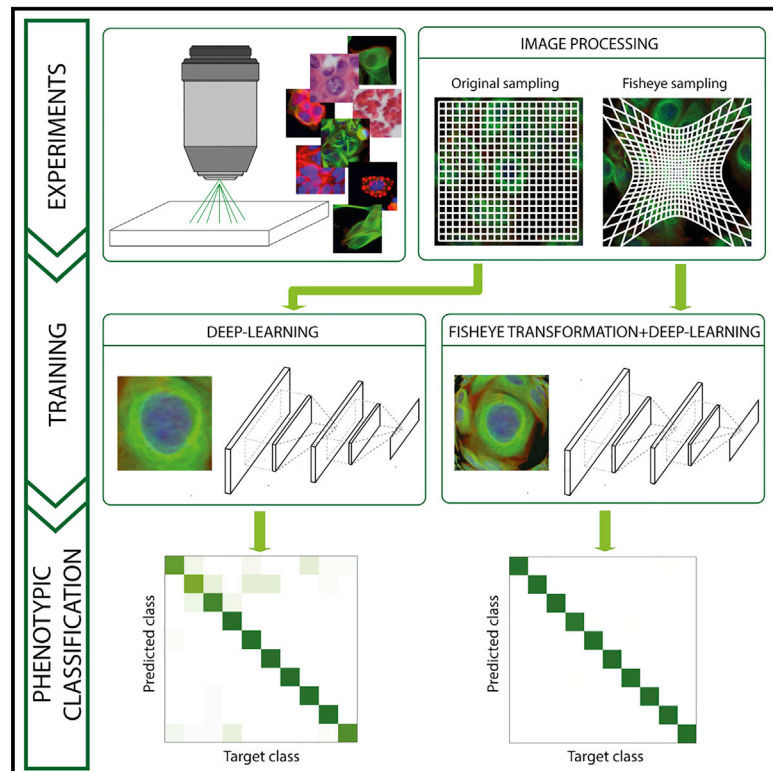
Authors

Timea Toth, David Bauer, Farkas Sukosd, Peter Horvath

Correspondence

horvath.peter@brc.hu

In brief

Toth et al. present an imaging approach to improve phenotypic classification accuracy of cells by taking into account the surrounding microenvironment of the cell. The method uses a transformation similar to that used in fisheye cameras and combines it with deep-learning algorithms to show increased accuracy for phenotypic classification of cells in both cell culture and tissue section samples.

Highlights

- Accurate cell classification is important in image-based cell phenotyping

- Accounting for the microenvironment of cells improves this accuracy

- Fisheye-like transformation increases accuracy with deep learning

- The results can have significant biological and clinical implications

CelPress

Cell Reports Methods

Report

Fisheye transformation enhances deep-learning-based single-cell phenotyping by including cellular microenvironment

Timea Toth,[1,2] David Bauer,[1] Farkas Sukosd,[3] and Peter Horvath[1,4,5,6,*]
[1]Synthetic and Systems Biology Unit, Biological Research Centre, Eötvös Loránd Research Network, Szeged, Hungary
[2]Doctoral School of Biology, University of Szeged, Szeged, Hungary
[3]Department of Pathology, University of Szeged, Szeged, Hungary
[4]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
[5]Single-Cell Technologies, Inc., Szeged, Hungary
[6]Lead contact
*Correspondence: horvath.peter@brc.hu
https://doi.org/10.1016/j.crmeth.2022.100339

MOTIVATION Samples are increasingly being examined at the single-cell level in biological and clinical research. A growing and powerful set of approaches to do so involve image-based phenotyping using machine-learning algorithms. Improving the accuracy of cell classification in such methods is of utmost importance and an ongoing goal in the field. Here, we address this need by developing an approach that takes into account the surrounding environment of a cell of interest and utilizes a fisheye-like transformation, along with the deep-learning algorithm, to improve accuracy of cell classification in both cell culture and tissue-based images.

SUMMARY

Incorporating information about the surroundings can have a significant impact on successfully determining the class of an object. This is of particular interest when determining the phenotypes of cells, for example, in the context of high-throughput screens. We hypothesized that an ideal approach would consider the fully featured view of the cell of interest, include its neighboring microenvironment, and give lesser weight to cells that are far from the cell of interest. To satisfy these criteria, we present an approach with a transformation similar to those characteristic of fisheye cameras. Using this transformation with proper settings, we could significantly increase the accuracy of single-cell phenotyping, both in the case of cell culture and tissue-based microscopy images, and we present improved results on a dataset containing images of wild animals.

INTRODUCTION

All living entities adjust to their environment, manifested as visually observable morphological differences both at the macro- and microscales. Therefore, incorporating microenvironmental information into object classification may have an enormous impact on the accuracy of evaluation.[1] This phenomenon may be of particular interest when the phenotypes of cells are determined.

The modern technological advancements in microscopy, sequential hybridization,[2,3] and mass spectrometry[4] have paved the way to evaluate cellular structures at high spatial and temporal resolution. These measurements generate large datasets; hence, automated computational methods are required to obtain objective information from these images.[5,6] Utilizing the inherent potentials of automated image analysis offers several advantages: it eliminates operator bias, provides quantitative data, and identifies visual characteristics that would otherwise go undetected.[7–9]

Various computer vision and classical machine-learning techniques have been used[10] to support researchers with tasks like image exploration (e.g., to find changes in cell structure in an imaging-based drug screen[11]), image classification (e.g., to determine the distribution of different proteins within cells[12]), image segmentation (e.g., identifying single cells in images[13]), or object tracking.[14] Despite the acknowledged capabilities of these techniques, deep-learning-based analyses often perform more efficiently in recognizing biological patterns based on the pixels of images.[15,16]

Deep learning has yielded fascinating results in solving biology-related issues.[17] The phenotype of a cell is determined by various cellular processes and factors (including the

stochasticity of gene expression, as well as a variety of proteomes and metabolomes[4]) that result in a particular morphological arrangement. Deep learning has enabled the exploration of factors like replicative age, organelle inheritance, and response to stress.[18] It has been demonstrated to perform comparably to human pathologists upon classifying whole-slide images into two categories of cancerous and normal lung tissues (it was even able to predict the ten most commonly mutated genes).[19] Another frequent challenge in cell biology lies in identifying various proteins and determining their locations within the cells. Numerous models have been developed[20,21] to automatically identify subcellular localization patterns, based on the Human Protein Atlas,[22] which contains acquisitions of 12,003 human proteins at the single-cell level.

Single-cell heterogeneity within cell populations is also influenced by the cell's microenvironment.[23,24] Several studies have demonstrated that the peculiarities of cellular neighborhood can be exceptionally relevant when investigating the collective organization of cells in a variety of settings. Snijder et al. have reported that in a cell culture context, one may predict the burden of viral infection at the single-cell level solely based on each cell's microenvironment.[24] In a study of competitive interactions between wild-type Madin-Darby canine kidney (MDCK) cells and cells depleted of the polarity protein scribble, Bove et al. found that the probability of cell division is significantly higher for MDCK cells when their neighborhood is mostly populated by scribble cells.[25] Several other examples of the importance of cellular neighborhood are also published in literature. For instance, neighboring epidermal stem cells affect each other (differentiation of a single stem cell is followed by division of a direct neighbor);[26] ligand-producing hair cell precursors in the inner ear are smaller than their neighbors.[27]

In a previous study, our group has also concluded that incorporating the features of the microenvironment of cells improves phenotype classification in high content screens.[1] In that study, we extracted commonly used cell-based features for every segmented cell. The center of mass was measured for each segmented area and was used as a reference point for distance calculation. We used two different approaches to define cellular neighborhoods: the K-nearest neighbors (KNN) and the N-distance methods (Figure 1A). Neighborhood features were derived from the mean, median, SD minimum, and maximum statistics of the previously calculated cellular features. Then, we used these neighborhood features to classify cells, and we got the best result using the multi-layer perceptron classifier. Based on these findings, we hypothesized that it is worth using environmental data for deep-learning phenotypic profiling.

Recently, fisheye cameras have received significant interest from both technical professionals and the public in general. Fisheye lenses are ultra-wide-angle lenses capable of taking wide panoramic or hemispheric images; however, they incorporate a significant optical distortion into the process. Fisheye lenses utilize specific mapping (stereographic, equidistant, equisolid angle, orthogonal), which provides the images a characteristic convex non-rectilinear appearance.[28] These specific lenses have a wide range of applications due to their ability to provide rich visual information, including the genera-

tion of augmented or virtual reality,[29] improving the performance of intelligent robot vision systems,[30] and simplifying the complexity of surveillance systems.[31] Various correction models have been proposed to rectify the distortion of fisheye lenses.[32–34]

In this report, we introduce a way of representing images to deep-learning-based image classification networks. The basic idea is the following: the original image includes the object of interest (which is located in the middle of the image), as well as its microenvironment of a pre-defined range. The images are then transformed by a fisheye-like spatial sampling method that collects more pixels from the close proximity of the object of interest, and the resolution decreases for larger proximity (Figure 1D). Our results indicate that the proposed transformation highly outperforms classical machine-learning methods and deep-learning-based classifiers benchmarked on cell cultures, scans of cancerous tissues, and real-life images. We remark that our fisheye transform method provides improved classification accuracy, as demonstrated by higher accuracy scores on several test datasets, compared with relying on feeding the network with multi-scale images in parallel (i.e., an image pyramid). Also, the presented fisheye transformation has the advantage that it can be incorporated into the network as a layer,[35] although technically it is more resource intensive when large images are fed into the network.

RESULTS

We evaluated the performance of a fisheye-like sampling on several image datasets (Figure 2), aiming to improve classification accuracy by deep-learning-based image classification networks. We compared two convolutional neural-network-based classifiers using ResNet50 and InceptionV3 backbones, as well as the relevance of the extent of neighborhood and the focal distance of fisheye transformation. We benchmarked what combination of these parameters produce the best results.

The baselines we compared our method with were also calculated with ResNet50 and InceptionV3 models. In the case of the cell culture and tissue section datasets, the input images were generated as follows: based on the center of the cells, we cropped out 192 × 192 pixel-sized images around them, then resized these images to 224 × 224 to meet the requirement of the models. For the iWildCam dataset, we cropped out the inside of the bounding boxes that were drawn around the animals, then resized these to 224 × 224 pixel images. For benchmarking, we have not used the fisheye transformation. For the statistical analysis of the classification accuracy results, two-sample t tests were performed (STAR Methods).

Based on previous studies in cellular biology, we expected that taking the cell's microenvironment into consideration improves the performance of deep-learning classification. To test this hypothesis, we have introduced a fisheye transformation, as this kind of distortion considers more pixels from the direct neighborhood of the object of interest than from the region beyond that (Figure 1D). We also hypothesized that there should be an optimal neighborhood range and focal distance combination with respect to deep-learning-based classification accuracy.
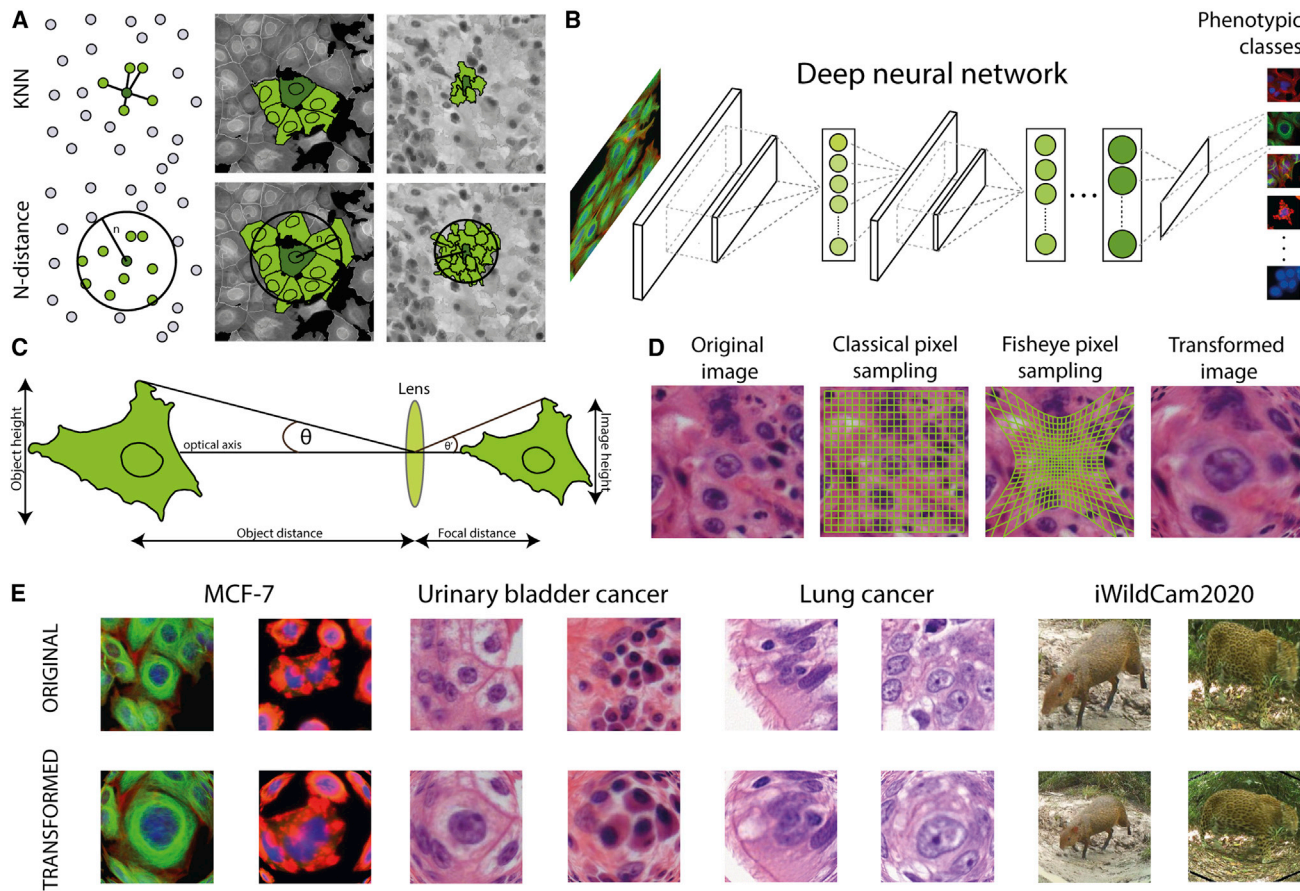
Figure 1. Phenotypic classification and fisheye transformation

(A) Neighborhood with classical machine learning. The K-nearest neighbors (KNN) and N-distance methods, illustrated in a schematic figure and in real cell culture and tissue section scenarios (K = 5, cell culture: n = 19.51 μm, tissue sections: n = 13.5 μm).

(B) Schematic figure of phenotypic classification with deep learning.

(C) Illustration of the optical parameters for the fisheye transformation.

(D) The difference between classical and fisheye pixel sampling: in the classic case, we select pixels evenly, while in the case of fisheye, sampling is dense near the object of interest and less dense as the distance from the object increases.

(E) Examples of the fisheye transformation.

Increased classification accuracy on images of cell cultures

In case of the MCF-7 breast cancer cell database, an environmental range of 45–724 pixels (17.56–282.54 μm) was defined. For comparison, the average nuclei size in this dataset is 37 pixels (14.44 μm). Validation results indicate that applying the fisheye transformation improves the accuracy of both (Resnet50 and InceptionV3) classifiers (Figure 3A). The best performance was achieved when we applied a window size of 543 pixels (211.91 μm) with a focal length of 130 arbitrary units (where 1 unit corresponds to the size of a pixel), using the ResNet50 model. In this case, accuracy reached 91.38%, which is 7% better than that achieved when using deep learning only (84.31%). The highest classification accuracy achieved with fisheye distortion also outperformed our previous results with classical machine-learning approaches,[1] where the maximum accuracy was 90.80% with the support vector machine classifier. Although the deep-learning baseline was higher for InceptionV3 (85.85%)

than for ResNet50, the best result we could achieve with InceptionV3 using distorted images was only 89.33%.

Looking at the fisheye-transformed images, it is reasonable to wonder whether we can achieve similarly high results by using only the nuclei to train our deep-learning network. In the case of the MCF-7 dataset, we did nuclei and cytoplasm segmentation (as described in STAR Methods). To demonstrate that the increase in classification accuracy is due to the inclusion of the environment and the fisheye transformation, we ran calculations on images showing only the nucleus or the nucleus and cytoplasm of the cell. We used the same ResNet50 deep-learning network as before for the baseline and fisheye calculations (we have gotten the best results with this network). The results show the same tendency that we expected: that the classification accuracy is lowest when the network sees only the nuclei and is best when we use the fisheye-transformed images and the network sees the cell's microenvironment (see Figures S5 and S6).
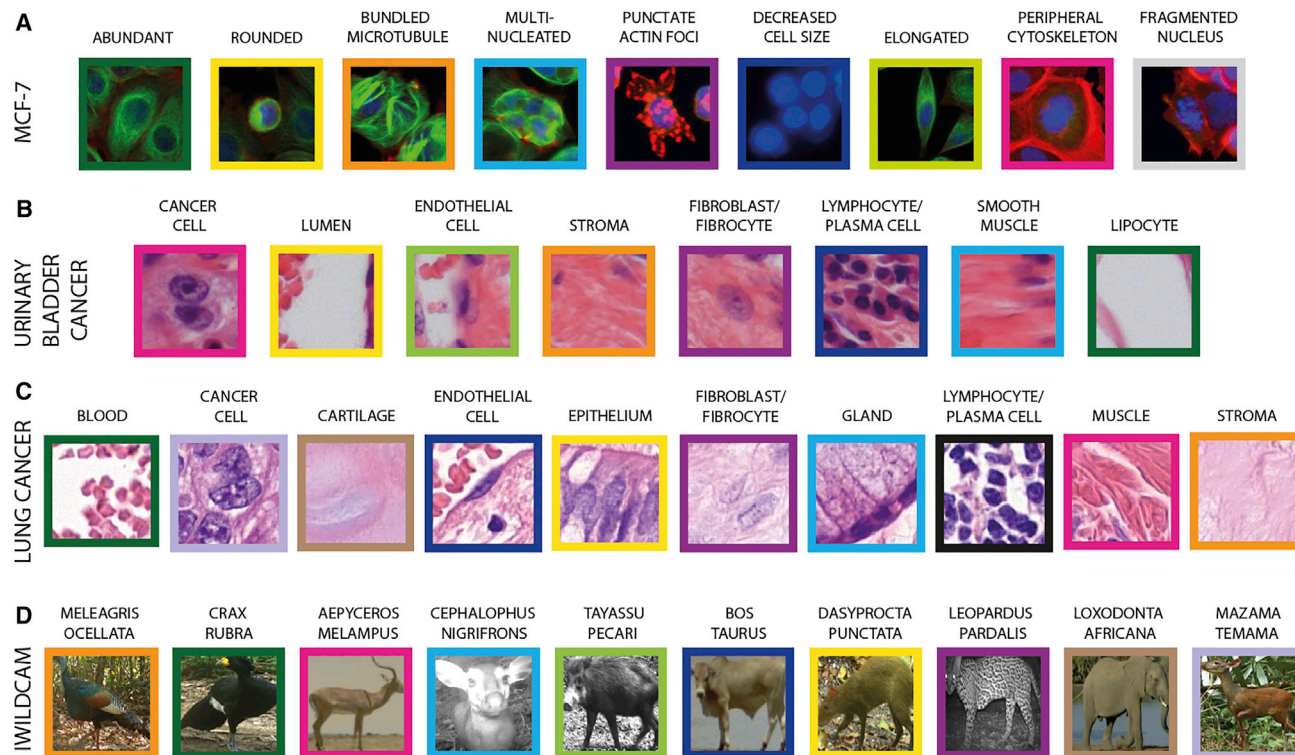
Figure 2. Distinguished classes
(A) Cells of nine different phenotype classes identified in the MCF-7 high-content-screening dataset.
(B) Eight phenotypic classes in the UBC tissue image dataset.
(C) Ten phenotypic classes in the LC tissue image dataset.
(D) The ten most common animal species in the iWildCam2020 dataset.

Fisheye transformation has a major impact on phenotyping tissue sections

In the case of the urinary bladder cancer (UBC) tissue image dataset, we also collected neighborhood information for a 45 to 724 window size (12.15–195.48 μm), where the average nuclei size was 32 pixels (8.64 μm), while for the lung cancer dataset, we used a range of 45–362 pixels (17.55–141.18 μm) with an average nuclei size of 39 pixels (15.21 μm). For both datasets, we were able to achieve higher classification accuracy on the fisheye-transformed images than with traditional deep learning, irrespective of whether we used ResNet50 or InceptionV3.

For the UBC dataset, we had previous results using neighborhood features with classical machine learning.[1] Then, the maximum of classification accuracy reached 93.37% with multi-layer perception (MLP) calculations. In our current study, the best performance reached 98.14%, appearing at a window size of 272 pixels with 150 arbitrary unit focal length using ResNet50 (Figure 3B). Without using the distorted images, classification accuracy was only 94.41%. Using InceptionV3, both the deep-learning baseline (93.24%) and the highest accuracy achieved on fisheye-distorted images (97.65%) were less favorable than those yielded with ResNet50.

The lung cancer dataset is the only exception where InceptionV3 performed slightly better than ResNet50 (Figure 3C). With ResNet50, the highest classification accuracy was 99.36%, while InceptionV3 yielded a maximum of 99.46% upon incorporating the neighborhood feature with a 272 pixel range and using 170 arbitrary units as focal distance. This is more than 2% better than the results yielded with InceptionV3 on undistorted images (accuracy: 97.25%).

Fisheye transformation outperforms image pyramids

In order to benchmark our proposed fisheye transformation against classical multi-scale approaches, we tested the phenotypic classification accuracy on UBC and lung cancer (LC) datasets using an image pyramid as an input to the networks. For this purpose, we fed the ResNet50 networks with images of varying (1/1, 1/2, 1/4) scales in parallel. Accuracy reached 97.9% and 99.1% for the UBC and LC datasets, respectively, indicating that applying the image pyramid approach yields better results than traditional deep learning but performs less favorably compared with the fisheye-distorted solution.

Improved accuracy in the case of the iWildCam2020 dataset

We investigated whether the inclusion of the environment of real scenes can improve classification results. The dataset included photos of animals taken from fixed camera positions. Because of the perspective, the animals could appear either very small or large in the images. To handle this sort of discrepancy, we
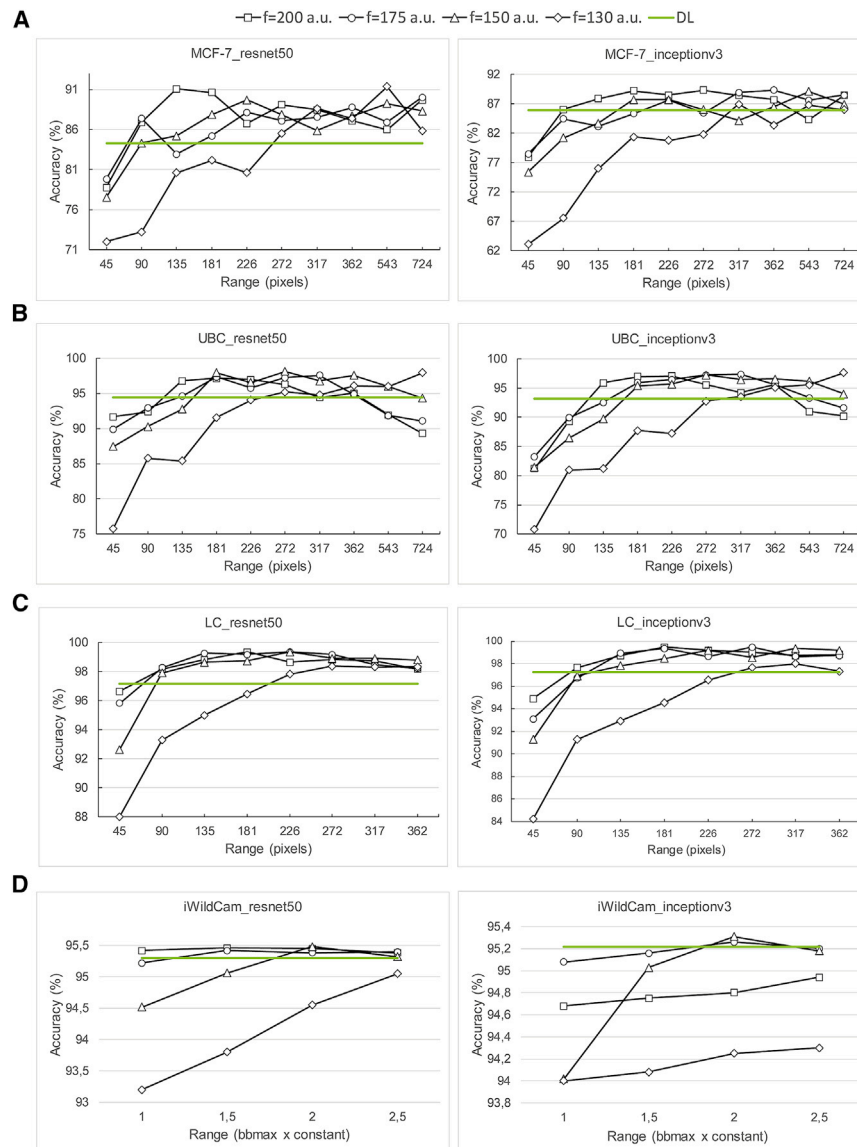
Figure 3. Comparison of the performance of deep-learning networks (ResNet50, InceptionV3) upon considering different neighborhood distances

(A) Classification accuracies for the MCF-7 cell culture dataset using ResNet50 (left) and InceptionV3 (right).

(B) Classification accuracies for the UBC tissue image dataset.

(C) Classification accuracies for the LC tissue dataset.

(D) Classification accuracies for the iWildCam2020 dataset. Green lines indicate the baseline yielded with deep learning upon using the original (undistorted) images, while black lines indicate the results achieved on fisheye-distorted images with different f (focal distance) values (the values are measured in arbitrary units).

and neighborhood features. Our results were compared with the accuracies obtained with deep learning (ResNet50 and InceptionV3), where the inputs for the net were non-fisheye-transformed images. For all four datasets we used (a cell culture, two tissue sections, and a dataset containing images of animals), training with fisheye-transformed images resulted in significantly higher accuracies (Figure 4). Though our method is robust, it is easy and fast to use. Furthermore, the fisheye-distortion approach is generally applicable to any kind of image data, where the environment has an influential role, as demonstrated by applying this method on the iWildCam2020 dataset.

The highest improvement in accuracy appeared when we applied our method on tissue section images. Generally, micro-environmental differences are visible in tissue histology studies. These are obvious manifestations of the cooperation and interdependence of different cells, which are also characterized histoanatomically. For example, the robust adjacency information of endothelial cells is the presence of a lumen on one side of them or the presence of well-known restricted cell types, such as connective tissue cells and smooth muscle cells, on the other side.

In the case of cell cultures, we also achieved an improvement in phenotyping accuracy. The explanation relies on two factors. Firstly, homogeneous-looking areas do not consist of molecularly completely identical cells.[24] This is captured by the fisheye transformation similarly as described in case of cancer tissues. Secondly, in homogeneous cellular regions, the neighborhood may provide a statistically more stable and, consequently, more powerful basis for decision.

We observed minor improvement in accuracy in the case of iWildCam data. A possible explanation is that during the recording of this dataset, fixed positioned cameras were used.

considered the size of the bounding boxes around the animals as references instead of using fixed pixel distances. The baselines for classification accuracy with traditional deep learning were 95.3% and 95.22% upon using ResNet50 and InceptionV3, respectively. In the case of the fisheye-transformed images, classification accuracy reached 95.48% with ResNet50 when 2.5× the size of bounding boxes was considered as the neighborhood feature and focal length was set to 150 units.
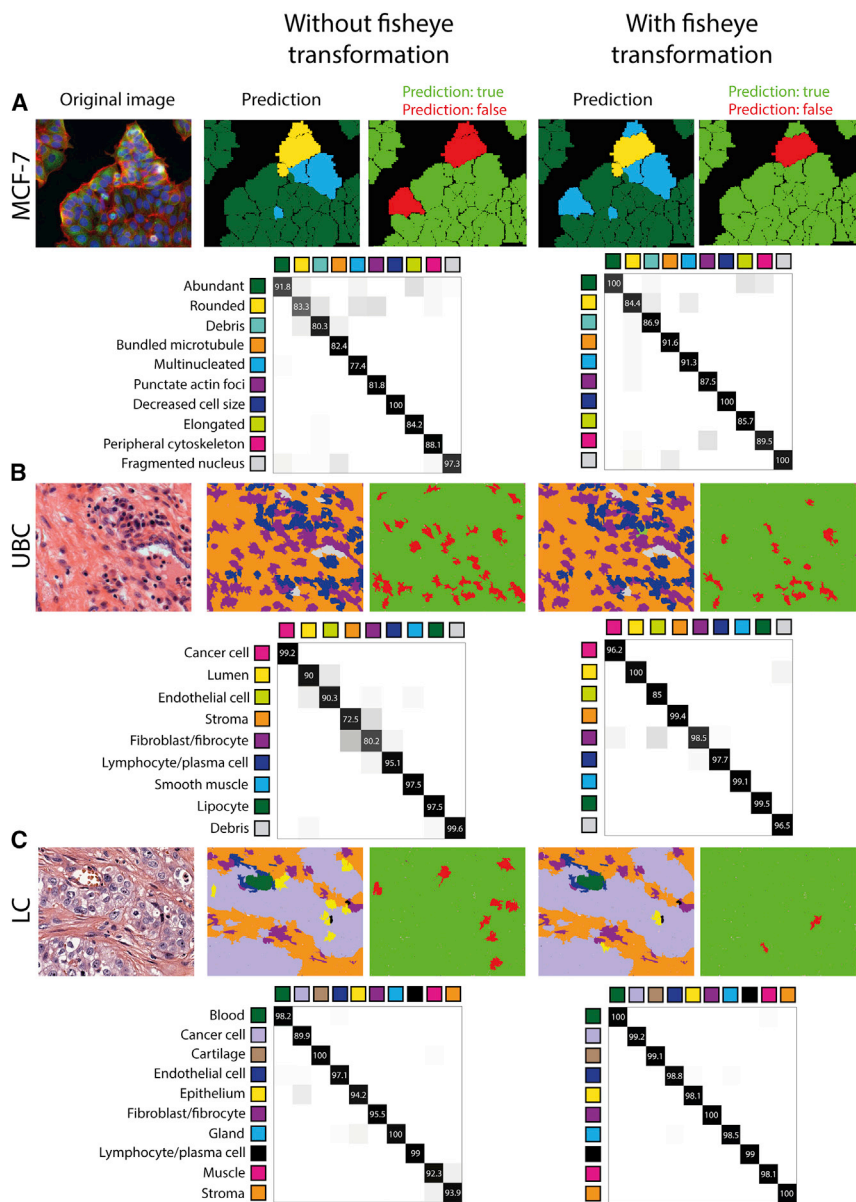
DISCUSSION

Here, we present a method combining fisheye transformation with deep learning, an extension to our previous model, incorporating the information obtained from the cellular microenvironment in phenotypic classification. We demonstrated on MCF-7 cell culture and UBC datasets that our method outperforms the approach of using classical machine learning with single-cell

Figure 4. The effect of combining the fisheye transformation with deep learning

(A) Prediction examples and confusion matrices based on ResNet50 models in the cell culture dataset. Original image (left), prediction and confusion matrix using the model built on standard images (middle), and prediction and confusion matrix using the model built on fisheye-transformed images, window size: 543 pixels, focal length: 130 arbitrary units (a.u.) (right). In the second and fourth images, we show with different colors the predicted phenotypes, and in the third and fifth images, we show if the prediction was true (green) or false (red).

(B) Prediction examples and confusion matrices of the best deep-learning performance (ResNet50) in the UBC tissue dataset. Original image (left), prediction and confusion matrix using traditional deep learning (middle), and considering the cellular neighborhood with fisheye transformation, window size: 272 pixels, focal length: 150 a.u. (right).

(C) Prediction examples and confusion matrices based on InceptionV3 models in the LC tissue dataset. Original image (left), prediction and confusion matrix using the model built on undistorted images (middle), and prediction and confusion matrix using the combination of deep-learning and fisheye-distorted images, window size: 272 pixels, focal length: 170 a.u. (right).

layers that are capable of learning non-linear spatial sampling functions.

Limitations of the study

In this work, we used transfer learning with two pre-trained networks. Because deep-learning systems learn gradually, they require huge amounts of data to train them. Creating a training set can be challenging as less data are available for rare phenotypes, making it difficult to obtain the amount of data needed to run the model correctly. Another limitation of our approach is that the proposed transformation is not integrated into the deep-learning network, so it implies an extra pre-processing step. During the deep-learning process, basic data augmentation has a seminal role in improving accuracy. And though it is possible to use standard geometry transformations (such as reflection, rotation, and scaling), it has a limitation, as any other transformation that spatially moves the cell of interest away from the center cannot be easily applied because that would have consequences to the fisheye transformation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

Because of this, depending on whether the animals are positioned near or far from the camera, one can see larger, smaller, or no surroundings. Therefore, even by increasing the window size, it may occur that we do not gain more information about the environment.

In conclusion, we show that the incorporation of the microenvironment into machine-based decisions can improve the task of classifying single cells into phenotypic classes. This confirms the fact that cellular structures are not arbitrarily organized and that it is beneficial to take these macro structures into consideration. We also show that using a non-uniform sampling of the original image data for deep-learning training and inference is feasible and can further improve accuracy. A potential extension to the presented approach could rely on the introduction of data transformer

- ● RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- ● EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Urinary bladder cancer and lung cancer tissue sections
- ● METHOD DETAILS
  - ○ Processing of datasets
  - ○ Segmentation
  - ○ Fisheye transformation
  - ○ Deep learning-based object classification
- ● QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Significance tests
  - ○ MCF-7
  - ○ Urinary bladder cancer
  - ○ Lung cancer
  - ○ iWildCam2020

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100339.

AUTHOR CONTRIBUTIONS

P.H. conceived and led the project. F.S. co-supervised the project. T.T. designed the pipeline. D.B. implemented the fisheye transformations in Python. T.T. prepared the figures. T.T., P.H., and F.S. designed the experiments and analyzed the data. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

P.H. is the founder and a shareholder of Single-Cell Technologies, Ltd.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Toth, T., Balassa, T., Bara, N., Kovacs, F., and Kriston, A. (2018). Environmental properties of cells improve machine learning-based phenotype recognition accuracy. Sci. Rep., 1–9. https://doi.org/10.1038/s41598-018-28482-y.

2. Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science 361. https://doi.org/10.1126/science.aat5691.

3. Zhu, Q., Shah, S., Dries, R., Cai, L., and Yuan, G.-C. (2018). Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. Nat. Biotechnol. 36, 1183–1190. https://doi.org/10.1038/nbt.4260.

4. Standke, S.J., Colby, D.H., Bensen, R.C., Burgett, A.W.G., and Yang, Z. (2019). Mass spectrometry measurement of single suspended cells using a combined cell manipulation system and a single-probe device. Anal. Chem. 91, 1738–1742. https://doi.org/10.1021/acs.analchem.8b05774.

5. Lock, J.G., and Stromblad, S. (2010). Systems microscopy: an emerging strategy for the life sciences. Exp. Cell Res. 316, 1438–1444. https://doi.org/10.1016/j.yexcr.2010.04.001.

6. Meijering, E., Carpenter, A.E., Peng, H., Hamprecht, F.A., and Olivo-Marin, J.-C. (2016). Imagining the future of bioimage analysis. Nat. Biotechnol. 34, 1250–1255. https://doi.org/10.1038/nbt.3722.

7. Caicedo, J.C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A.S., Barry, J.D., Bansal, H.S., Kraus, O., et al. (2017). Data-analysis strategies for image-based cell profiling. Nat. Methods 14, 849–863. https://doi.org/10.1038/nmeth.4397.

8. Dufour, A.C., Jonker, A.H., and Olivo-Marin, J.C. (2017). Deciphering tissue morphodynamics using bioimage informatics. Philos. Trans. R. Soc. B Biol. Sci. 372. https://doi.org/10.1098/rstb.2015.0512.

9. Keller, P.J. (2013). Imaging morphogenesis: technological advances and biological insights. Science 340, 1234168. https://doi.org/10.1126/science.1234168.

10. Grys, B.T., Lo, D.S., Sahin, N., Kraus, O.Z., Morris, Q., Boone, C., and Andrews, B.J. (2016). Machine learning and computer vision approaches for phenotypic profiling. J. Cell Biol. 216, 65–71. https://doi.org/10.1083/jcb.201610026.

11. Scheeder, C., Heigwer, F., and Boutros, M. (2018). Machine learning and image-based profiling in drug discovery. Curr. Opin. Syst. Biol. 10, 43–52. https://doi.org/10.1016/j.coisb.2018.05.004.

12. Lin, D., Sun, L., Toh, K.-A., Zhang, J.B., and Lin, Z. (2018). Biomedical image classification based on a cascade of an SVM with a reject option and subspace analysis. Comput. Biol. Med. 96, 128–140. https://doi.org/10.1016/j.compbiomed.2018.03.005.

13. Molnar, C., Jermyn, I.H., Kato, Z., Rahkama, V., Östling, P., Mikkonen, P., Pietiäinen, V., and Horvath, P. (2016). Accurate morphology preserving segmentation of overlapping cells based on active contours. Sci. Rep. 6, 32412. https://doi.org/10.1038/srep32412.

14. Meijering, E., Dzyubachyk, O., Smal, I., and van Cappellen, W.A. (2009). Tracking in cell and developmental biology. Semin. Cell Dev. Biol. 20, 894–902. https://doi.org/10.1016/j.semcdb.2009.07.004.

15. Pratapa, A., Doron, M., and Caicedo, J.C. (2021). Image-based cell phenotyping with deep learning. Curr. Opin. Chem. Biol. 65, 9–17. https://doi.org/10.1016/j.cbpa.2021.04.001.

16. Gupta, A., Harrison, P.J., Wieslander, H., Pielawski, N., Kartasalo, K., Partel, G., Solorzano, L., Suveer, A., Klemm, A.H., Spjuth, O., et al. (2019). Deep learning in image cytometry: a review. Cytom. Part A 95, 366–380. https://doi.org/10.1002/cyto.a.23701.

17. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. Nat. Methods 16, 1233–1246. https://doi.org/10.1038/s41592-019-0403-1.

18. Mattiazzi Usaj, M., Sahin, N., Friesen, H., Pons, C., Usaj, M., Masinas, M.P.D., Shuteriqi, E., Shkurin, A., Aloy, P., Morris, Q., et al. (2020). Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability. Mol. Syst. Biol. 16, e9243. https://doi.org/10.15252/msb.20199243.

19. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., and Tsirigos, A. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat. Med. 24, 1559–1567. https://doi.org/10.1038/s41591-018-0177-5.

20. Sullivan, D.P., Winsnes, C.F., Åkesson, L., Hjelmare, M., Wiking, M., Schutten, R., Campbell, L., Leifsson, H., Rhodes, S., Nordgren, A., et al. (2018). Deep learning is combined with massive-scale citizen science to

improve large-scale image classification. Nat. Biotechnol. *36*, 820–828. https://doi.org/10.1038/nbt.4225.

21. Ouyang, W., Winsnes, C.F., Hjelmare, M., Cesnik, A.J., Åkesson, L., Xu, H., Sullivan, D.P., Dai, S., Lan, J., Jinmo, P., et al. (2019). Analysis of the human protein Atlas image classification competition. Nat. Methods *16*, 1254–1261. https://doi.org/10.1038/s41592-019-0658-6.

22. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321. https://doi.org/10.1126/science.aal3321.

23. Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226. https://doi.org/10.1016/j.cell.2008.09.050.

24. Snijder, B., Sacher, R., Ramo, P., Damm, E.-M., Liberali, P., and Pelkmans, L. (2009). Population context determines cell-to-cell variability in endocytosis and virus infection. Nature *461*, 520–523. https://doi.org/10.1038/nature08282.

25. Bove, A., Gradeci, D., Fujita, Y., Banerjee, S., Charras, G., and Lowe, A.R. (2017). Local cellular neighborhood controls proliferation in cell competition. Mol. Biol. Cell *28*, 3215–3228. https://doi.org/10.1091/mbc.e17-06-0368.

26. Mesa, K.R., Kawaguchi, K., Cockburn, K., Gonzalez, D., Boucher, J., Xin, T., Klein, A.M., and Greco, V. (2018). Homeostatic epidermal stem cell self-renewal is driven by local differentiation. Cell Stem Cell *23*, 677–686.e4. https://doi.org/10.1016/j.stem.2018.09.005.

27. Shaya, O., Binshtok, U., Hersch, M., Rivkin, D., Weinreb, S., Amir-Zilberstein, L., Khamaisi, B., Oppenheim, O., Desai, R.A., Goodyear, R.J., et al. (2017). Cell-cell contact area affects notch signaling and notch-dependent patterning. Dev. Cell *40*, 505–511.e6. https://doi.org/10.1016/j.devcel.2017.02.009.

28. Sahin, C. (2017). The geometry and usage of the supplementary fisheye lenses in smartphones. Smartphones Appl. Res. Perspect. https://doi.org/10.5772/intechopen.69691.

29. Schmalstieg, D., and Hollerer, T. (2016). Augmented Reality: Principles and Practice (Addison-Wesley Professional).

30. Sáez, Á., Bergasa, L.M., López-Guillén, E., Romera, E., Tradacete, M., Gómez-Huélamo, C., and del Egido, J. (2019). Real-time semantic segmentation for fisheye urban driving images based on ERFNet. Sensors *19*. https://doi.org/10.3390/s19030503.

31. Tseng, D., Chen, C., and Tseng, C. (2017). Automatic detection and tracking in multi-fisheye cameras surveillance system. Int. J. Comput. Elect. Eng. *9*, 370–383. https://doi.org/10.17706/ijcee.2017.9.1.370-383.

32. Li, T., Tong, G., Tang, H., Li, B., and Chen, B. (2020). FisheyeDet: a self-study and contour-based object detector in fisheye images. IEEE Access *8*, 71739–71751. https://doi.org/10.1109/ACCESS.2020.2987868.

33. Silberstein, S., Levi, D., Kogan, V., and Gazit, R. (2014). Vision-based pedestrian detection for rear-view cameras. In 2014 IEEE Intelligent Vehicles Symposium Proceedings (IEEE), pp. 853–860. https://doi.org/10.1109/IVS.2014.6856399.

34. Bertozzi, M., Castangia, L., Cattani, S., Prioletti, A., and Versari, P. (2015). 360° detection and tracking algorithm of both pedestrian and vehicle using fisheye images. In 2015 IEEE Intelligent Vehicles Symposium (IV) (IEEE), pp. 132–137. https://doi.org/10.1109/IVS.2015.7225675.

35. Jaderberg, M., Simonyan, K., Zisserman, A., and kavukcuoglu, koray (2015). Spatial transformer networks. In Advances in Neural Information Processing Systems, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds. (Curran Associates, Inc.).

36. Caie, P.D., Walls, R.E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., Roberts, M.E., and Carragher, N.O. (2010). High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol. Cancer Ther. *9*, 1913–1926.

37. Piccinini, F., Balassa, T., Szkalisity, A., Molnar, C., Paavolainen, L., Kujala, K., Buzas, K., Sarazova, M., Pietiainen, V., Kutay, U., et al. (2017). Advanced cell classifier: user-friendly machine-learning-based software for discovering phenotypes in high-content imaging data. Cell Syst. *4*, 651–655.e5. https://doi.org/10.1016/j.cels.2017.05.012.

38. Mund, A., Coscia, F., Hollandi, R., Kovács, F., Kriston, A., Brunner, A.-D., Bzorek, M., Naimy, S., Rahbek Gjerdrum, L.M., Dyring-Andersen, B., et al. (2021). AI-driven Deep Visual Proteomics defines cell identity and heterogeneity. Preprint at: bioRxiv. https://doi.org/10.1101/2021.01.25.427969.

39. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. *34*, 2274–2282. https://doi.org/10.1109/TPAMI.2012.120.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Urinary bladder and Lung cancer tissue	University of Szeged	Ethics: authorisation number: 5127, registration number: 17/2022-SZTE
Deposited data		
MCF-7 dataset	Ljosa et al., 2012	https://bbbc.broadinstitute.org/BBBC021/
UBC dataset	This paper	https://doi.org/10.6084/m9.figshare.21259287.v1
LC dataset	This paper	https://doi.org/10.6084/m9.figshare.21259422.v1
iWildCam 2020 dataset	Wildlife Conservation Society, iNaturalist, the U.S. Geological Survey, and Microsoft AI for Earth	https://www.kaggle.com/competitions/iwildcam-2020-fgvc7/data
Phenotypic classes and coordinates of cells	This paper	https://doi.org/10.6084/m9.figshare.21259464.v1
Software and algorithms		
Python	Python Software Foundation	V3.8.10
Matlab 2019b	Mathworks, USA	https://www.mathworks.com/products/matlab.html
BIAS	Single-Cell Technologies	http://single-cell-technologies.com/
Fisheye transformation	This paper	https://doi.org/10.6084/m9.figshare.21259398.v1
Transfer learning for fisheye transformed images	This paper	https://doi.org/10.6084/m9.figshare.21259416.v1

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Peter Horvath (horvath.peter@brc.hu).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- This paper analyzes original and already existing data. All data used to generate the figure of this study are publicly available as of the date of publication and are listed in the key resources table.
- All original code is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in the paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Urinary bladder cancer and lung cancer tissue sections
We used images of urinary bladder cancer (UBC) and lung cancer (LC) tissues as test datasets (Figures 2B and 2C). The images were acquired of patient derived samples obtained from University of Szeged (approval authorisation number: 5127, registration number: 17/2022-SZTE). For the bladder cancer dataset, the samples obtained from 3 patients (sex and gender identity/age/stage of cancer of the subjects respectively: female/71/stage 2, male/65/stage 3, male/73/stage 2). For the lung cancer dataset, the samples obtained from 4 patients (sex and gender identity/age/histological observations of the subjects respectively: female/75/primary pulmonary multinodular invasive papillary adenocarcinoma, female/65/primary lung origin, acinar predominant adenocarcinoma, female/74/primary lung origin, invasive adenocarcinoma, acinaris predominant, male/76/primary lung origin, invasive solid adenocarcinoma).

The slides of urinary and lung cancer tissues were stained with hematoxylin-eosin (HE) in standard histopathological procedures. Formalin-fixed and paraffin-embedded tissue sections were cut into 4 μm thick slices, and were stained using a Tissue-Tek DRS 2000E-D2 Slide Stainer (Sakura Finetek Japan) according to the manufacturer's instructions. Using the AxioVision SE64Rel.4.9.1.1 (Carl Zeiss Meditec AG, Germany) software, images were captured with an Axio Imager Z.1 (Carl Zeiss Meditec AG, Germany) microscope equipped with an EC Plan-NEOFLUOAR 20x/0.5NA lens. All procedures were performed with the approval of the University of Szeged.

METHOD DETAILS

Processing of datasets

We used a publicly available breast cancer cell line (MCF-7) dataset, available online at the Broad Bioimage Benchmark Collection.[36] Using this dataset Piccinini and colleagues[37] published a single-cell phenotypic annotation method which we have adopted in our study. Nine phenotypic classes (i.e. abundant, rounded, elongated, multinucleated, bundled microtubule, peripheral cytoskeleton, punctate actin foci, decreased cell size and fragmented nucleus) and a debris class were identified (Figure 2A), and approximately 1,500 cells were labelled.

For the analysis of the bladder cancer tissue dataset, containing 38 images, we used the annotation presented in our previous neighbourhood study.[1] We distinguished eight phenotypic classes (cancer cell, lumen cell, endothelial cell, stroma cell, fibroblast-fibrocyte, lymphocyte-plasma cell, smooth muscle cell and lipocyte), and labelled 1,200 cells. For the lung cancer dataset, we differentiated 10 phenotypic classes (blood cell, cancer cell, cartilage, endothelial cell, epithelium, fibroblast-fibrocyte, gland, lymphocyte-plasma cell, muscle cell and stroma cell), and labelled 5,000 cells.

The iWildCam 2020 dataset is derived from the iWildCam 2020 competition organised by Kaggle. It consists of 217,959 images, where 267 classes were distinguished, containing a highly unbalanced number of entities, were defined. Figure 2D displays the 10 classes that contain most of the elements (excluding the 'empty' class, where no animals are visible in the image).

Segmentation

CellProfiler 2.2.0 was used to segment images from the high-content-screening dataset of drug-treated MCF-7 samples. The adaptive Otsu method was used to detect nuclei. Cells smaller than 5 μm and nuclei contacting the borders of the images were eliminated. Adaptive thresholding was used to extract the cytoplasm of cells, using watershed separation based on the nuclei as seed points.

For the segmentation of the tissue datasets and for the annotation of the MCF-7, UBC and LC datasets we used an image analysis and machine learning software named BIAS,[38] which was developed by Single-Cell Technologies Ltd. (Szeged, Hungary).

The SLIC superpixel segmentation technique[39] was employed to segment urinary bladder cancer and lung cancer section images. We set 35 pixel as superpixel size, and forced connectivity between superpixels when a superpixel included less than 25 pixels. We have previously demonstrated that this superpixel size works best to represent the cellular structure of tissues.[1]

During the annotation process of these three (MCF-7, urinary bladder and lung cancer) datasets, we saved the x-y coordinates of the centre of the nuclei/superpixels, and used these coordinates as inputs for the deep learning-based fisheye transformation.

For the iWildCam dataset, the organisers of the competition provided a general animal detection model called MegaDetector (https://github.com/microsoft/CameraTraps/blob/master/megadetector.md), along with an annotation file that contained one label per image. When more than one animal were visible in the image, we selected only the detection with the highest accuracy, and gave the label to it. MegaDetector works with bounding boxes. In the present study, we used the x-y coordinates of the centre of the bounding boxes as inputs for the fisheye transformation.

Fisheye transformation

Several types of ultra-wide angle lenses are available, and all are associated with a significant visual distortion. In our study we evaluated an algorithm that artificially reproduces the same kind of distortion which is inherent in images taken with ultra-wide angle lenses. We analysed the relevance of neighbourhood features to determine the most optimal distance to be considered for the highest accuracy of classification.
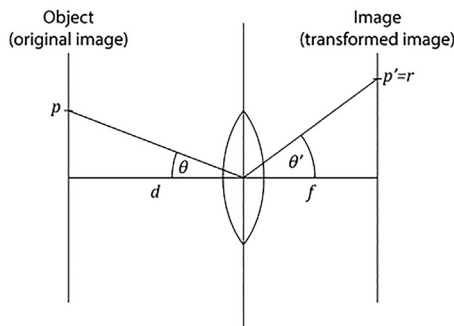
In fisheye transformation, the position of the projection of a given real world point can be determined from the angle of the incident ray. This can be calculated using a mapping function:

$$r = m(f, \theta),$$

where f is the focal length. For a given f we can reformulate the above equation as

$$r = m_f(\theta).$$

The mapping function is an inherent function to fisheye lenses. It defines the position of the object from the centre of the image (r) as a function of the focal distance (f) and the angle from the optical axis (θ). The functions in wide-angle lens cameras include the following:

- Rectilinear:

$$r = f \tan \theta$$

- Fisheye
  - Equidistant

$$r = f\theta$$

  - Equisolid angle

$$r = 2f \sin \frac{\theta}{2}$$

  - Stereographic

$$r = 2f \tan \frac{\theta}{2}$$

  - Orthographic

$$r = f \sin \theta$$

All of these functions are invertible:

$$\theta = m_f^{-1}(r).$$

The following relation is also valid for θ:

$$\tan\theta = \frac{p}{d},$$

where d is the distance between point p and the centre of the lens, measured along the axis of the lens. In our case the object points are the pixels of an image, and d is constant across the whole image.

In image transformation tasks the transformation function is usually given as an inverse mapping, which provides the source position of each output pixel, in this case, the value of p for each output position r. From the equations above, p can be calculated as follows:

$$p = d \cdot \tan m_f^{-1}(r) \,.$$

This equation contains two free parameters: d and f. The value of d appears as a scalar multiplier, hence it affects magnification only. The value of f affects both the scale and the magnitude of distortion. However, in our case we endeavoured to change the strength of the fisheye effect of the transformation only, leaving the scale of the selected area unaffected. In order to achieve the

desired effect, it is possible to connect the value of d to the value of f in such a way that it unaffects scaling.
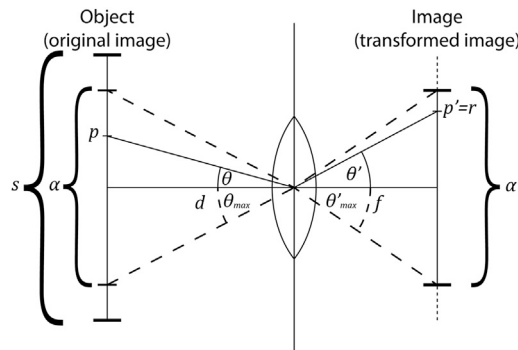
Let's mark the size of the selected area with α. We wish to keep the position of the corner points intact while applying the fisheye distortion. In this case an equation is introduced as follows:

$$tan\theta_{max} = \frac{\alpha}{2d},$$

where θ_{max} is the angle of the incoming ray from the borders of the selected area. As for the border points are expected to be transformed into themselves,

$$\theta_{max} = m_f^{-1}\left(\frac{\alpha}{2}\right)$$

is also valid. Combining these two equations gives

$$d = \frac{\alpha}{2tan\ m_f^{-1}\left(\frac{\alpha}{2}\right)}.$$

It is evident that the last equation is determined by f as a modifiable parameter (when the value of α is fixed). Thus, the final form of the equation for the fisheye transformation is

$$p = \frac{\alpha}{2}\cdot\frac{tan\ m_f^{-1}(r)}{tan\ m_f^{-1}\left(\frac{s}{2}\right)}.$$

As an easy-to-read interpretation, we normalise the value of θ to the range of [0, 1], and then rescale it to the range $\left[0,\frac{\alpha}{2}\right]$.

Note that in the calculations above it is assumed that the selected area of interest is in the middle of the image. However, with a simple translation, the calculations are valid for any arbitrary image positions.

The first parameter we optimised was the range around our object-of-interest (in an optical manner, this is object height, later referred as 'window size', see Figure 1C). For the MCF-7, UBC and LC datasets, as we have mentioned before, we saved the x-y coordinates of the centre of the nuclei/superpixels and then we used the selected pixel range for the fisheye transformation based on these points. In case of the iWildCam dataset, we cropped out areas of interest from the original images for our fisheye transformation in four different sizes. We took the size of the original bounding box and multiplied it with 1.0, 1.5, 2.0 and 2.5. When we multiplied the original size with 1.0, we cropped the images with the same size for the baseline images and for the fisheye-transformation. In all of the other cases, we took a bigger size of the environment into account.

The second modifiable parameter was focal distance, which essentially defines the strength of the distortion. Contrary to our original neighbourhood idea,[1] information acquisition was not cell-based, but pixel-based (Figures 1A and 1D). The third parameter was the mapping function, which originally (in cameras) is responsible for transforming a part of a spherical object to a 2D plane. In the present study we chose the 'equidistant' function -as it is one of the most popular mapping functions used in cameras - to test our hypothesis on the significance of neighbourhood regarding classification accuracy. In all cases we set the object distance (the distance between the original image and the lenses) in a way to eliminate scaling due to fisheye distortion. For more examples of the transformed images, see Figures S1–S4.

Deep learning-based object classification

For image classification we used Matlab R2019b and its Deep Learning Toolbox (version 13.0). This toolbox provides a framework to design or implement networks, pre-trained models, and apps. Two networks pre-trained on ImageNet, ResNet50 and InceptionV3 were utilised for transfer learning. Our decision for the models (i.e. using ResNet50 and InceptionV3), was based on the accuracy, speed and size of the networks. Particular attention was paid to avoid that an annotated cell is included in the training dataset and appears in a validation image as a neighbour afterwards. Thus, to prevent a potentially positive influence of evaluation, both the MCF-7 and UBC datasets were processed with pre-defined train and validation folders at the image level, instead of the cell level.

For the LC and iWildCam datasets this was not an issue, as these were created to contain only one annotated object per image. The latter two datasets were randomised and split, thus 80% was used for training and 20% was used for validation. We used only images not previously seen by the network for validation. Three independent trials were run for each dataset.

Data augmentation has a seminal role in improving accuracy. We applied the following standard geometry transformations on our training datasets: reflection in the left-right and a top-bottom directions, rotation, horizontal and vertical scaling. It is important to mention that we have not used any transformation that spatially moves the cell-of-interest away from the centre of the image because that would have consequences to the fisheye transformation.

The deep-learning parameters used: MiniBatchSize: 64, MaxEpochs: 100, InitialLearnRate: 3e-4, LearnRateDropFactor: 0.3, LearnRateDropPeriod: 50, Shuffle: every epoch.

QUANTIFICATION AND STATISTICAL ANALYSIS

Significance tests
As a statistical procedure, two-sample t-test was performed, we considered the result significant at $p < 0.05$. For detailed results see Table S1.

MCF-7
Our best result (accuracy: 91.38%) for this dataset with the fisheye-transformation appears with a window size of 543 pixels with a focal length of 130 arbitrary units when we used ResNet50. We compare this result to a baseline, where we used ResNet50 (accuracy: 84.31%). For the baseline we cropped out images around the cells' centre with 192 × 192 pixel diameter (so in this case, we haven't performed fisheye transformation on the original images). Deep learning calculations were run 5–5 times on both the baseline and the fisheye transformed data. At the 0.05 level, the 7.07% difference in accuracy is significant.

Urinary bladder cancer
Our best result (accuracy: 98.14%)for this dataset with the fisheye-transformation appears with a window size of 272 pixels with a focal length of 150 arbitrary units when we used ResNet50. We compare this result to a baseline, where we used ResNet50 (accuracy: 94.41%). For the baseline we cropped out images around the cells' centre with 192 × 192 pixel diameter (so in this case, we haven't performed fisheye transformation on the original images). Deep learning calculations were run 5–5 times on both the baseline and the fisheye transformed data. At the 0.05 level, the 3.83% difference in accuracy is significant.

Lung cancer
Our best result (accuracy: 99.46%)for this dataset with the fisheye-transformation appears with a window size of 272 pixels with a focal length of 170 arbitrary units when we used inceptionV3. We compare this result to a baseline, where we used inceptionV3 (accuracy: 97.25%). For the baseline we cropped out images around the cells' centre with 192 × 192 pixel diameter (so in this case, we haven't performed fisheye transformation on the original images). Deep learning calculations were run 5–5 times on both the baseline and the fisheye transformed data. At the 0.05 level, the 2.21% difference in accuracy is significant.

iWildCam2020
Our best result (accuracy: 95.48%)for this dataset with the fisheye-transformation appears when 2.5× the size of bounding boxes were considered as the neighbourhood feature and focal length was set to 150 units and we used ResNet50. We compare this result to a baseline, where we used ResNet50 (accuracy: 95.3%). For the baseline we cropped out images with different dimensions based on the bounding boxes provided by the Kaggle organisers (so in this case, we haven't performed fisheye transformation on the images). Deep learning calculations were run 5–5 times on both the baseline and the fisheye transformed data. At the 0.05 level, the 0.18% difference in accuracy is significant.