



Quantification of Osteoclasts in Culture, Powered by Machine Learning

OPEN ACCESS

Edited by:

Fernando Antunes,
University of Lisbon, Portugal

Reviewed by:

Helen Knowles,
University of Oxford, United Kingdom
Teun J. De Vries,
VU University Amsterdam,
Netherlands
Hugo Alexandre Ferreira,
University of Lisbon, Portugal

*Correspondence:

Yankel Gabet
yankel@tauex.tau.ac.il
Drorit Neumann
histo6@tauex.tau.ac.il
Amir Globerson
gamir@tauex.tau.ac.il

† These authors have contributed
equally to this work and share first
authorship

‡ These authors have contributed
equally to this work and share senior
authorship

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 01 March 2021

Accepted: 26 April 2021

Published: 25 May 2021

Citation:

Cohen-Karlik E, Awida Z,
Bergman A, Eshed S, Nestor O,
Kadashev M, Yosef SB, Saed H,
Mansour Y, Globerson A, Neumann D
and Gabet Y (2021) Quantification
of Osteoclasts in Culture, Powered by
Machine Learning.
Front. Cell Dev. Biol. 9:674710.
doi: 10.3389/fcell.2021.674710

Edo Cohen-Karlik^{1†}, Zamzam Awida^{2†}, Ayelet Bergman¹, Shahar Eshed¹, Omer Nestor¹, Michelle Kadashev², Sapir Ben Yosef², Hussam Saed², Yishay Mansour¹, Amir Globerson^{1*}, Drorit Neumann^{2} and Yankel Gabet^{3*‡}**

¹ Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel, ² Department of Cell and Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, ³ Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

In vitro osteoclastogenesis is a central assay in bone biology to study the effect of genetic and pharmacologic cues on the differentiation of bone resorbing osteoclasts. To date, identification of TRAP+ multinucleated cells and measurements of osteoclast number and surface rely on a manual tracing requiring specially trained lab personnel. This task is tedious, time-consuming, and prone to operator bias. Here, we propose to replace this laborious manual task with a completely automatic process using algorithms developed for computer vision. To this end, we manually annotated full cultures by contouring each cell, and trained a machine learning algorithm to detect and classify cells into *preosteoclast* (TRAP+ cells with 1–2 nuclei), *osteoclast type I* (cells with more than 3 nuclei and less than 15 nuclei), and *osteoclast type II* (cells with more than 15 nuclei). The training usually requires thousands of annotated samples and we developed an approach to minimize this requirement. Our novel strategy was to train the algorithm by working at “patch-level” instead of on the full culture, thus amplifying by >20-fold the number of patches to train on. To assess the accuracy of our algorithm, we asked whether our model measures osteoclast number and area at least as well as any two trained human annotators. The results indicated that for osteoclast type I cells, our new model achieves a Pearson correlation (*r*) of 0.916 to 0.951 with human annotators in the estimation of osteoclast number, and 0.773 to 0.879 for estimating the osteoclast area. Because the correlation between 3 different trained annotators ranged between 0.948 and 0.958 for the cell count and between 0.915 and 0.936 for the area, we can conclude that our trained model is in good agreement with trained lab personnel, with a correlation that is similar to inter-annotator correlation. Automation of osteoclast culture quantification is a useful labor-saving and unbiased technique, and we suggest that a similar machine-learning approach may prove beneficial for other morphometrical analyses.

Keywords: osteoclasts, automatic quantification of osteoclasts, machine learning, object detection, deep learning, convolutional neural network (CNN), deep neural networks (DNN), artificial intelligence

INTRODUCTION

Bone is a highly dynamic tissue that undergoes continuous remodeling throughout life, in a process involving the concerted actions of monocyte-derived osteoclasts that resorb mineralized tissue, and mesenchymal osteoblasts that deposit new bone (Teitelbaum, 2007; Clarke, 2008; Karsenty et al., 2009).

Bone mass is carefully maintained by a tight coordination of the activity of the bone-resorbing osteoclasts and bone-forming osteoblasts (Boyle et al., 2003; Teitelbaum and Ross, 2003; Takayanagi, 2007; Zaidi, 2007), and an imbalance between their activities results in skeletal pathologies such as osteoporosis (Bruzzaniti and Baron, 2006; Novack and Teitelbaum, 2008). Compounds demonstrating any ability to modulate the balance between osteoclasts and osteoblasts are thus of great interest in treating such diseases.

Osteoclast precursor differentiation to functionally active multinucleated osteoclasts depends on administration of macrophage colony stimulating factor (M-CSF) and receptor activator for nuclear factor kappa B ligand (RANKL) (Teitelbaum, 2000).

This process, called osteoclastogenesis can be examined *in vitro* using a well-established and commonly used assay in which bone marrow derived macrophages are cultured with M-CSF and RANKL (Marino et al., 2014). During differentiation, the cells acquire a higher expression of tartrate-resistant acid phosphatase (TRAP) (Minkin, 1982) and fuse together to become multinucleated cells (Hata et al., 1992). Osteoclasts are commonly defined *in vitro* as TRAP positive (using specific staining) multinucleated cells. This assay is important for the experimental screening of therapeutic candidates targeting osteoclastogenesis and bone resorption.

The current gold standard method for quantifying osteoclast formation in culture is based on manual counting of TRAP-positive, multinucleated (≥ 3 nuclei) cells visualized under the microscope, which is both a subjective and a time-consuming process of evaluation (Marino et al., 2014). To overcome this significant drawback, we now present an unbiased high-throughput approach that enables fast and accurate measurement of osteoclast number and area for more efficient screening of potential therapeutic agents in bone biology. With the development of machine learning techniques, image classification and object detection applications are becoming more accurate and robust. As a result, machine learning based methods are being applied in a wide range of fields. The use of computer vision algorithms for the analysis of microscopic images has received growing interest in recent years. Such methods have been used for tasks such as detecting cell membranes (Ciresan et al., 2012), detecting cell nuclei (Xue and Ray, 2017), and segmenting cells (Ronneberger et al., 2015). This success motivated us to develop such an approach for osteoclasts. We note that existing tools cannot be directly applied to our setting due to its unique characteristics (see below).

Here, we report the development of an artificial intelligence-based object detection method designed to identify, classify, and quantify osteoclasts in cultures.

To our knowledge, there is no publicly- or commercially available software dedicated to *in vitro* osteoclast detection and evaluation. Previously developed AI-based approaches for cell detection are not applicable for this purpose due to its unique characteristics; the structure of osteoclasts is different from other cells (multinucleated TRAP-stained cells); collecting data for training an AI system dedicated to osteoclasts is time consuming (i.e., annotating a single culture takes several hours to annotate), thus limiting the number of examples that can be obtained in a timely manner. In order to overcome these challenges, we describe a method we developed to train our model on TRAP-stained osteoclast cultures and thereby increase the effective size of the training data.

For the purpose of training and evaluating our algorithm, we manually annotated 11 full osteoclast cultures containing thousands of cells. Osteoclast precursor cultures were treated with M-CSF and RANKL to induce differentiation. Images resulting from the TRAP staining were used to train the system to identify and count multinucleated TRAP+ osteoclasts, and to further validate and generalize the method. The trained model was tested on images that were not included in the training phase.

MATERIALS AND METHODS

Materials

Minimum Essential Medium α (Alpha-MEM) and fetal bovine serum (FBS), referred to here as “Standard medium,” were purchased from Rhenium (Modiin, Israel), and culture plates were from Corning (New York, NY, United States). As a source of M-CSF, we used supernatant from CMG 14–12 cells, containing 1.3 $\mu\text{g/ml}$ M-CSF (Takeshita et al., 2000). RANKL was purchased from R&D Systems, Minneapolis, MN, United States.

Animals

Female wild type mice of the inbred strain C57BL/6J-RccHsd, aged 8–12 weeks were purchased from Envigo (Israel) and housed at the Tel-Aviv University animal facility. These mice were used for the generation of bone marrow derived macrophages (BMDM). Animal care and all procedures were in accordance with and with the approval of the Tel Aviv University Institutional Animal Care and Use Committee (Permit number 01-19 -032).

Cell Culture

Bone marrow cells were harvested from femurs and tibias of 8- to 12-week-old female mice. Cells were seeded on tissue-culture treated plates in standard medium (alpha-MEM supplemented with 10% fetal bovine serum). On the following day, non-adherent cells were seeded in non-tissue culture-treated plates in standard medium supplemented with 100 ng/ml M-CSF, which induces cell proliferation and differentiation into preosteoclasts.

For the osteoclastogenesis assay, preosteoclasts were plated in 96-well plates (8,000 cells per well) with standard medium supplemented with 20 ng/ml M-CSF and 50 ng/ml RANKL, replaced every 2 days. On the 4th day, cells were stained using a TRAP staining kit (Sigma-Aldrich, St. Louis, MO, United States),

and multinucleated (≥ 3 nuclei) TRAP-positive cells were defined as osteoclasts. Images were acquired at a magnification of $\times 4$ (Evos FLC, Life Technologies, MS, United States) (Hiram-Bab et al., 2015). An open-source graphical image annotation tool was used to measure osteoclast number and surface area from a single operator manual tracing for each well as previously described (Wada, 2016).

Machine Learning Pipeline

Machine learning algorithms are used in a wide range of domains and underlie many technologies such as speech recognition, image understanding, machine translation, fraud detection, and face recognition.

To understand how machine learning algorithms work, we can consider an image categorization problem where the task of the model is to label an image with the name of the object in the image. Formally, the goal is to map from an input image X to an output label Y . The learning model is then simply a function from X to Y . In order to learn this function, one collects a “training” dataset of X - Y pairs (i.e., images and the corresponding correct label), and the learning process seeks a function that fits this training data well.

A key question is of course what type of functions to use when mapping X to Y . In recent years, functions mimicking neural networks have been found to work particularly well for a wide range of problems (Figure 1). This approach is also known as deep learning because it involves a multi-layered computation process.

Arguably, the most striking successes of deep learning to date are in the field of computer vision, where the goal is to develop algorithms that perform a semantic analysis of images, in a manner similar to the human visual system. This field has undergone a revolution since the introduction of the AlexNet architecture (Krizhevsky et al., 2012), and many other more advanced architectures since (Simonyan and Zisserman, 2014; Szegedy et al., 2015; He et al., 2016). AlexNet is an example of a so-called Convolutional Neural Network (CNN) model. These models use specific connectivity patterns between layers to form an architecture that utilizes the spatial structure within images (Figure 2).

Object Detection

Our goal in this study was to automatically evaluate the number and area of osteoclasts in TRAP-stained cultures. To perform this task, we need to detect the cells in an image and then predict the total number of cells and the ratio between the area covered by cells and the size of the culture. This problem is closely related to the machine vision problem of *Object Detection*, which is designed to locate and correctly classify objects in images.

Object detection is a key task in computer vision and has been the focus of much research in recent years (Girshick et al., 2014; Girshick, 2015; Ren et al., 2015). Algorithms developed for this purpose are at the heart of many evolving technologies including autonomous vehicles (Menze and Geiger, 2015) and robotics (Jia et al., 2011). The purpose of object detection pipelines is to locate and correctly classify objects in images (see Figure 3). The architecture used for objective detection is a variant of the CNNs

described above, but with additional mechanisms for finding multiple objects in an image and providing suitable bounding boxes and visual categories.

This study is based on the Single Shot Detection (SSD) architecture (Liu et al., 2016), which has proven useful for detecting small objects (Lam et al., 2018). SSD consists of two parts: (1) A *backbone* Convolutional Neural Network that extracts features from an input image, and (2) several convolutional layers with detection heads that output bounding boxes and the corresponding class followed by Non-Maximum Suppression (NMS) to filter overlapping bounding boxes (Felzenszwalb et al., 2009; Girshick et al., 2014; Girshick, 2015; Figure 4).

The most common use of SSD is for object detection in natural images. As a training set for such SSD models, one uses large datasets of natural images that have been annotated with bounding boxes for visual objects, as well as their visual labels such as the Pascal VOC and MSCOCO datasets (Lin et al., 2014; Everingham et al., 2015).

Here our focus is quite different, as we are interested in detecting images in cell cultures, and thus models trained on natural images are not directly applicable. Instead, the approach we took here is to collect a new dataset of annotated cell images and use this to train a new model. This process will be described in more detail.

Data Collection

One of the main challenges in utilizing deep learning algorithms in various domains is the need for vast amounts of labeled data for training. Manually labeling data for object detection is a time consuming and expensive process due to the need to annotate each bounding box. In addition, the human annotator may require specific training. In this section we describe the data collection process and considerations.

For data annotation, we used an open-source annotation tool (Wada, 2016) to mark polygons around the cells (Figure 5). Marking polygons is more time-consuming than bounding boxes but it allows for a more versatile application such as running algorithms that quantify the contour of cells and segmentation (not tested in the scope of the current study).

We divided the cells into four different types (Figure 6): (1) *preosteoclast* – TRAP+ cells with 1–2 nuclei; (2) *osteoclast type I* – cells with more than (\geq) 3 nuclei and less than 15 nuclei; (3) *osteoclast type II* – cells with more than (\geq) 15 nuclei; (4) *ghost cells* – vanished cells which are distinguishable by their silhouette.

Full cultures may contain over 500 cells (Figure 5) and are hard to label accurately. To produce meaningful annotations, we divided each culture into equally sized regions, which are each annotated as separate images. This introduces a trade-off between the annotation speed and the precision of the annotations. We found that dividing each culture well to 16 regions, produced accurate annotations in a reasonable time.

For the annotation process, regions without any cells were discarded; this resulted in 133 regions with an average of 66.4 cells per region from 11 wells. Manually annotating a region took between several minutes for regions with few cells, and over an hour for regions with more than 100 cells (see Figure 7). Among

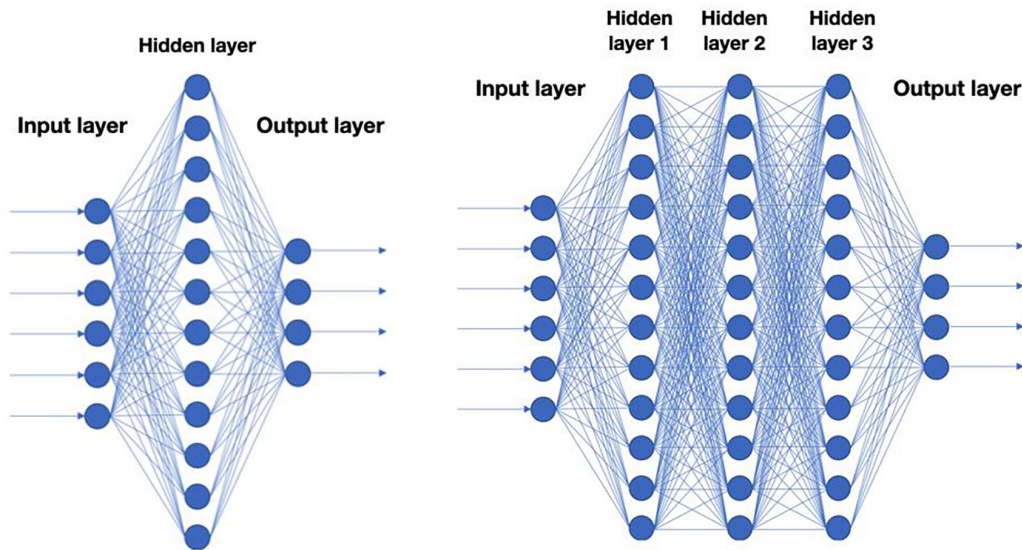


FIGURE 1 | Schematic representation of neural nets. Each circle corresponds to a *neuron* in a neural net. A neuron is typically a linear function of the previous layer followed by a non-linearity. Left: A neural net with a single hidden layer. Right: Deep neural net with 3 hidden layers.

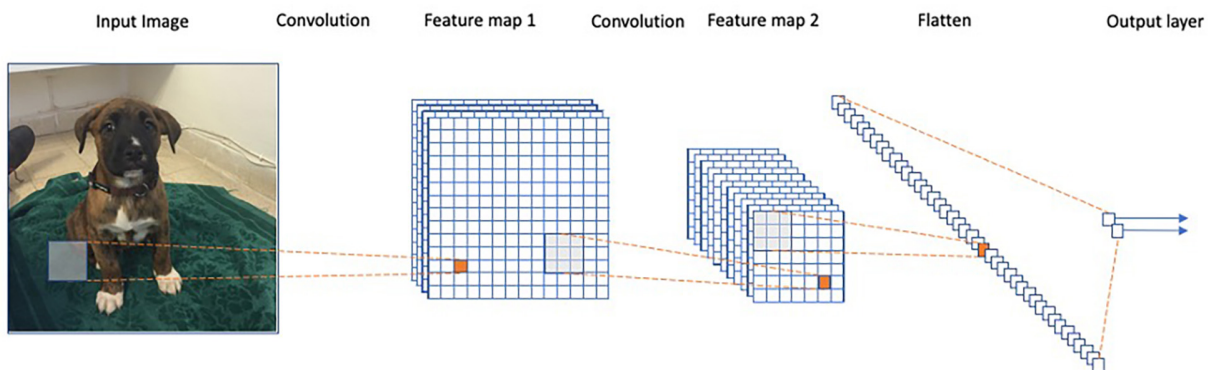


FIGURE 2 | Representation of a convolutional neural network. Each convolution operation consists of applying the same filters across the image to obtain a feature map in the following layer. The final iteration produces a fully connected layer similar to a regular neural net.

the 133 regions we annotated, 27 regions had more than 100 cells (**Figure 7** [right]).

Patch Generation Strategy

Deep learning algorithms typically require a large number of examples for effective training. The considerations and challenges in annotating cell cultures described in the section “Data Collection” restrict the number of fully labeled cultures available. Thus, our main challenge was to train a deep neural network on relatively few training samples. A major advance made here stems from the realization that, in contrast to “images in the wild” where a single object can span most of the image, images of cultured cells can be viewed at a wide range of resolutions and preserve their semantic properties due to the small size of each cell compared to the entire culture. For example, when estimating the number of cells and the area covered by cells, we can restrict

ourselves to small sections of the culture at any given point of time and still produce a perfectly annotated culture. This is not true when objects extend over a large section of the image such as in **Figure 3**.

Motivated by this observation, our strategy was to operate in small patch levels of $\frac{1}{8} \times \frac{1}{8}$, which means that each culture comprises 64 non-overlapping patches. This strategy also provides many different options for sampling (overlapping) patches and offers a substantial increase in the number of samples the model can be trained on. Specifically, we sampled a random region of the culture with 10 to 15% of the original width and height. The upper-left coordinate of the region defines the patch location. This process generates an enormous number of options for selecting a patch from a given culture. For example, given an image of a culture of size $1,000 \times 1,000$, there are more than $850^2 \cdot 50^2 (\approx 1.8 \cdot 10^9)$ different ways to sample a patch by this



FIGURE 3 | Output of SSD trained on Pascal VOC. The algorithm detects and wraps the objects with bounding boxes as well as correctly classifying them into the relevant class (the classification is depicted by the bounding box color: red corresponds to “Person” class, blue to “Dog” class, yellow to “Car” class, and green to “Bicycle” class).

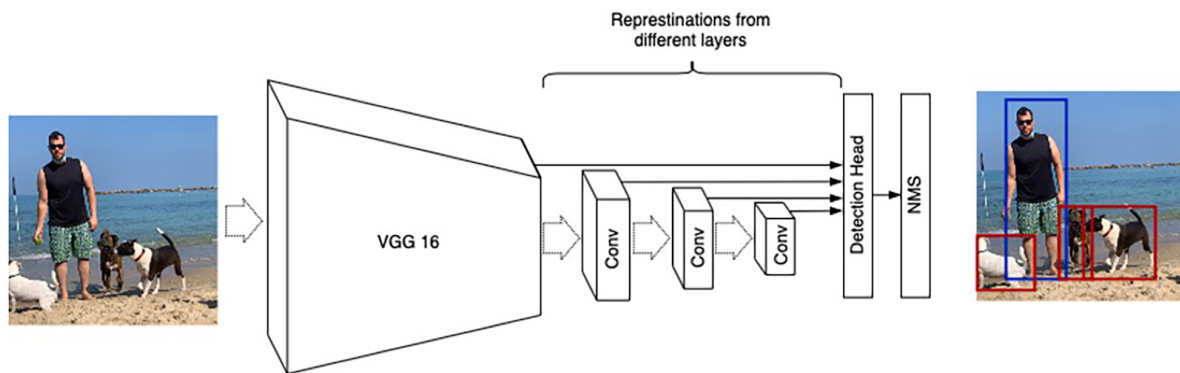


FIGURE 4 | SSD architecture illustration. VGG16 (Simonyan and Zisserman, 2014) is used as a base network. Detection heads are connected to numerous convolution layers (denoted as “Conv”) to account for different resolutions.

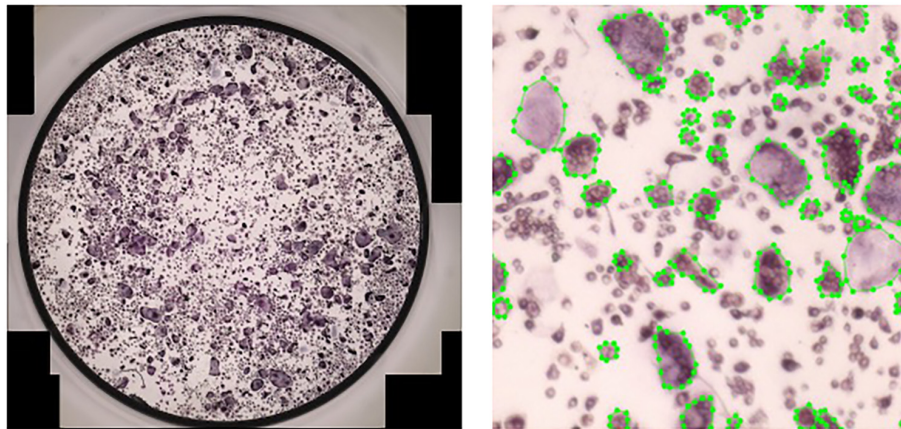


FIGURE 5 | Manual annotation of osteoclasts in cultures. Left: A full culture to be annotated. Right: A crop of size $\frac{1}{16} \times \frac{1}{16}$ of a fully annotated culture.

procedure. Here, we sampled 3,600,000 random patches from 10 wells used for training.

Another mechanism for increasing the size of training data, is to apply non-informative transformations to training points.

For example, adding a small amount of noise to an image is unlikely to change the semantic content, but provides a new data point for the algorithm to train on. In the deep-learning literature, this is known as “Data Augmentation”

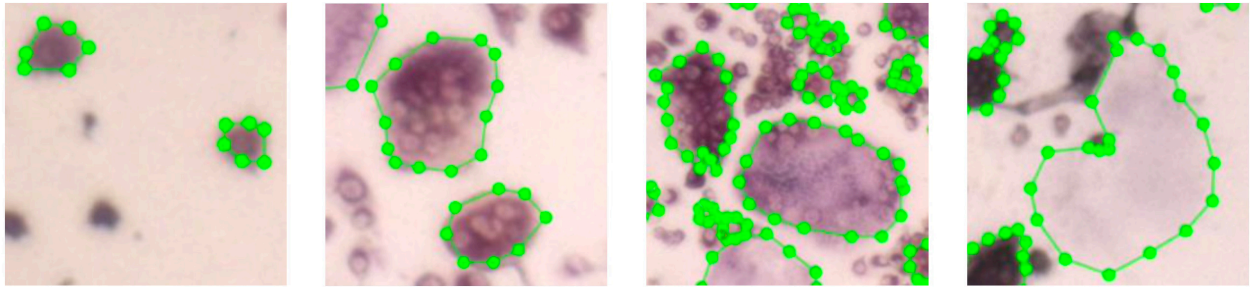


FIGURE 6 | Images of the four different cell types. The first three images depict cells with different number of nuclei. The rightmost cell is a ghost cell.

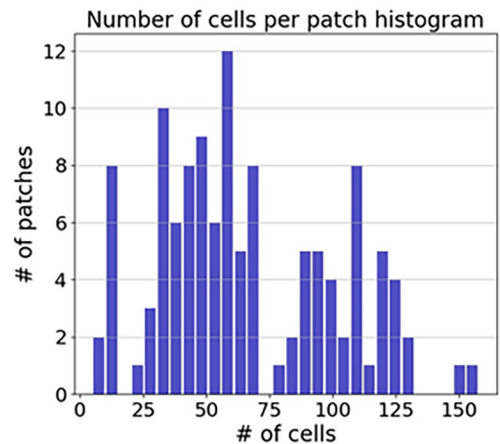
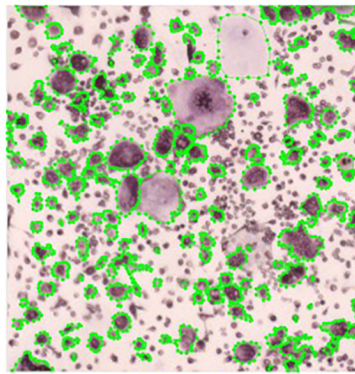
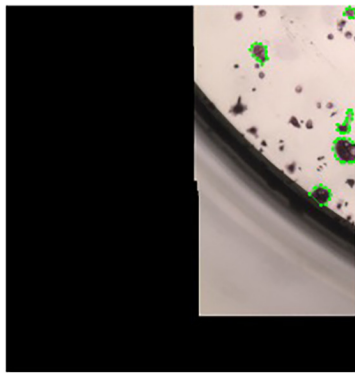


FIGURE 7 | Distribution of the number of cells per region. The number of cells per region ranged from 5 annotated cells (left) to 158 annotated cells (middle). The right panel shows a histogram representing the number of cells per region.

(Shorten and Khoshgoftaar, 2019). Here, we augmented the training data by applying a combination of the following transformations: (i) random photometric distortions, affecting brightness, hue, and saturation, (ii) random vertical flips of the patches, and (iii) rotations of the patches by 90° , 180° , and 270° .

Training the model with only 10 images of complete cultures did not allow for meaningful image detection. In contrast, the above procedures for generating random patches significantly increased the size of the training data, and thus improved the learned model.

Configuration and Hyperparameters

Our implementation was written in PyTorch, a deep learning library (Paszke et al., 2019) written in Python and C++. We used an open-source project as our starting point for SSD¹. Since object detection pipelines were designed to detect bounding boxes, we transformed each polygon created during the annotation process to the tightest square containing the entire cell (Figure 8).

The implementation of SSD contains more than a dozen different hyperparameters, and we adapted some settings to fit our unique data. The main change involves reducing the number of convolution layers from 6 to 4, which removes bounding

boxes spanning large sections of an input image. These objects are common in natural data (Figures 3, 4) and less relevant for culture well images (Figure 8). The modification of the architecture along with our strategy to operate on patches allows us to detect cells with sizes 51–383 μm .

Training

Because our scheme for augmentation extracts patches from full cultures, we dedicated 10 cultures for training, and 1 culture for testing. As is common practice for object detection, we utilized VGG16 (Simonyan and Zisserman, 2014), which was previously trained on ImageNet (Russakovsky et al., 2015) as the backbone network, and resumed training from these initial weights. We trained for 120,000 iterations with a batch size of 32 using an Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.00001.

Detection in Large Images

As already discussed, our model was trained to detect cells in small patches. We therefore also needed a mechanism that could allow us to apply the model to images that are larger than these small patches. A simple strategy, which we employed here, is to divide the large image into a set of non-overlapping

¹<https://github.com/amdegroot/ssd.pytorch>

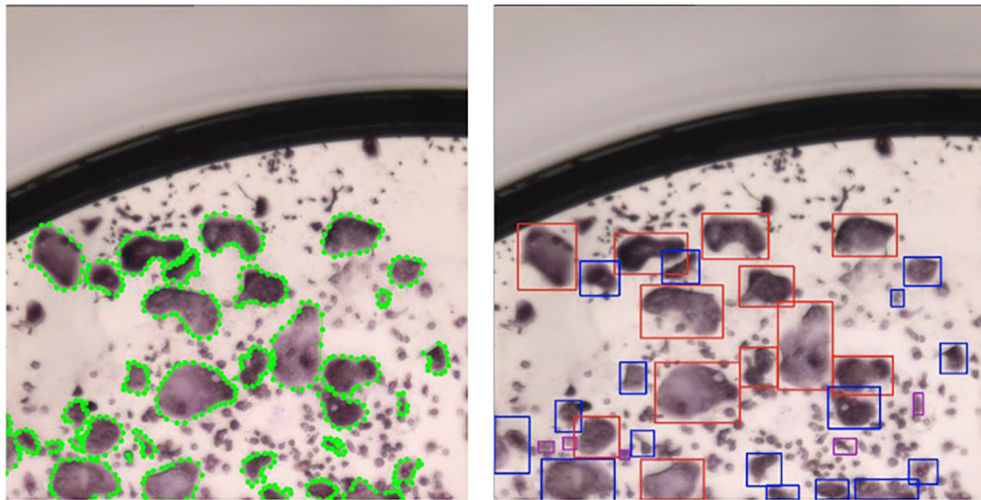


FIGURE 8 | SSD implementation using bounding boxes. Left: patch with polygons annotated by a human annotator; Right: patch after transforming the polygons to the tightest square surrounding a polygon. Blue bounding boxes correspond to type I osteoclasts and red bounding boxes correspond to type II osteoclasts.

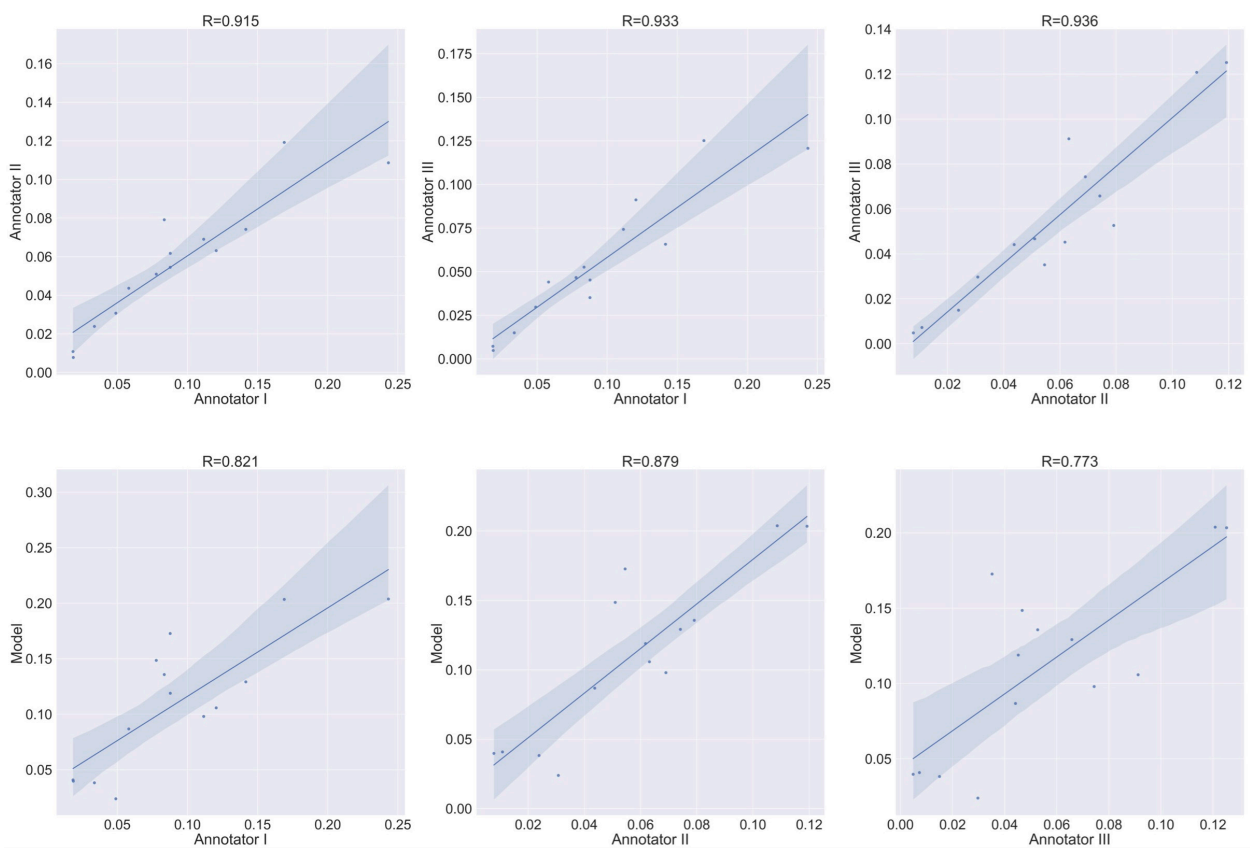


FIGURE 9 | Agreement level on the cell counting task of Type 1 osteoclasts. Shaded area represents the variation boundaries. Upper row: correlation between the three annotators for the cell counting task performed on the same set of culture images. Bottom row: correlation between each annotator and the model's prediction for the cell counting task.

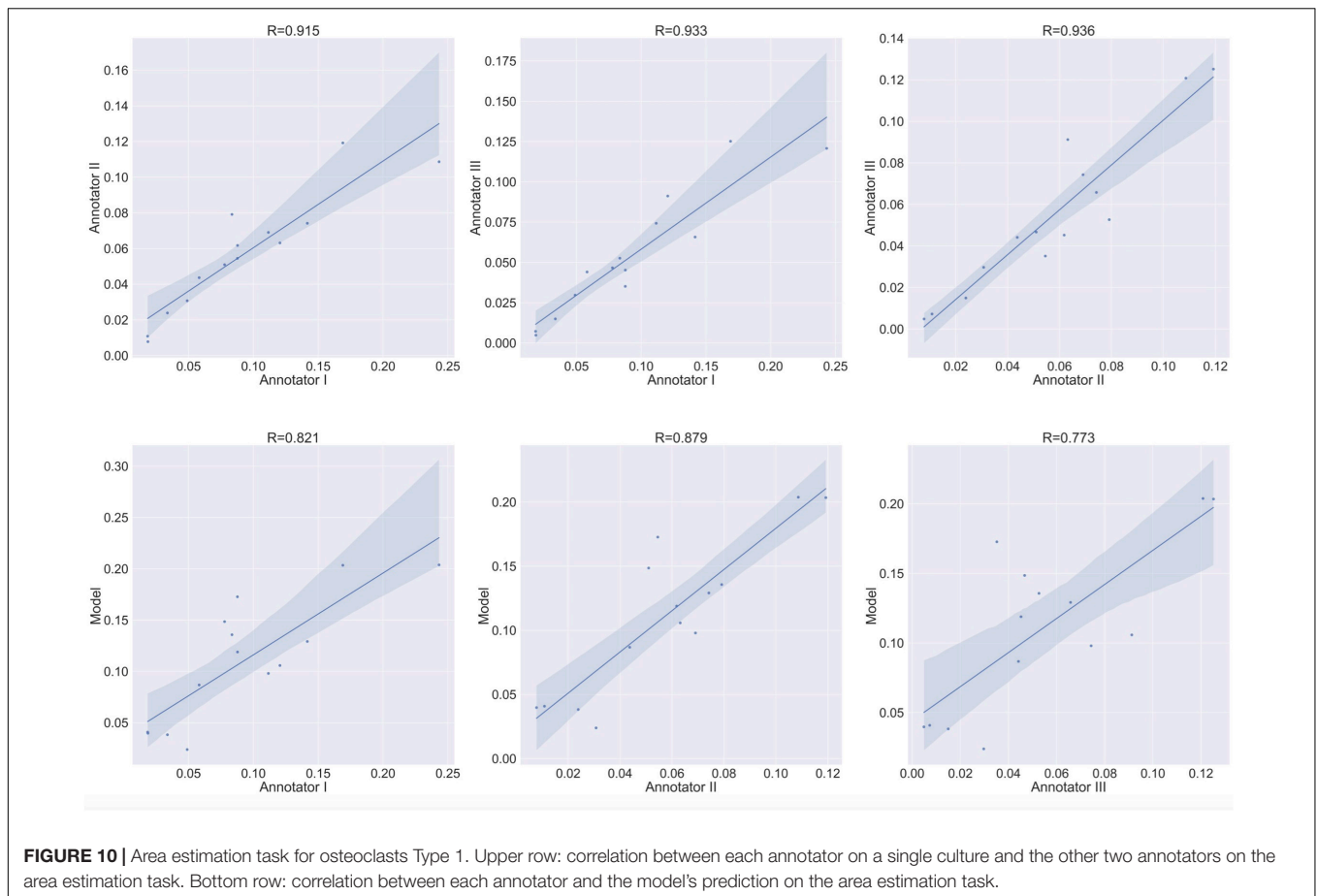


TABLE 1 | Correlation coefficients between human annotators and the model for cell counting.

	Annotator II	Annotator III	Model
Correlation coefficients for preosteoclast numbers			
Annotator I	0.506	0.819	0.879
Annotator II		0.687	0.730
Annotator III			0.920
Correlation coefficients for Type 2 osteoclast numbers			
Annotator I	0.952	0.955	0.968
Annotator II		0.960	0.913
Annotator III			0.921
Correlation coefficients for Ghost cell numbers			
Annotator I	0.597	0.287	0.269
Annotator II		0.781	0.410
Annotator III			0.423

TABLE 2 | Correlation coefficients between human annotators and the model for cell area measurements.

	Annotator II	Annotator III	Model
Correlation coefficients for preosteoclasts area			
Annotator I	0.656	0.865	0.937
Annotator II		0.708	0.753
Annotator III			0.891
Correlation coefficients for Type 2 osteoclasts area			
Annotator I	0.980	0.978	0.995
Annotator II		0.991	0.978
Annotator III			0.978
Correlation coefficients for Ghost cells area			
Annotator I	0.326	-0.007	0.590
Annotator II		0.917	0.300
Annotator III			0.070

small patches and apply the model to each one individually. The output is then simply the union of the outputs from the small patches. Since our model was trained on $\frac{1}{8} \times \frac{1}{8}$ images, we had 64 patches per culture. It is true that splitting images into non-overlapping patches may split some cells across multiple patches and may therefore be considered a disadvantage (see “Error Analysis”). However, in practice, we found that the naive

approach worked satisfactorily, and more sophisticated schemes were not necessary.

Evaluation

The main focus in bone cell culture analysis is to quantify the number and area of the cells rather than predicting the location of the cells inside the culture. Thus, the natural metrics for

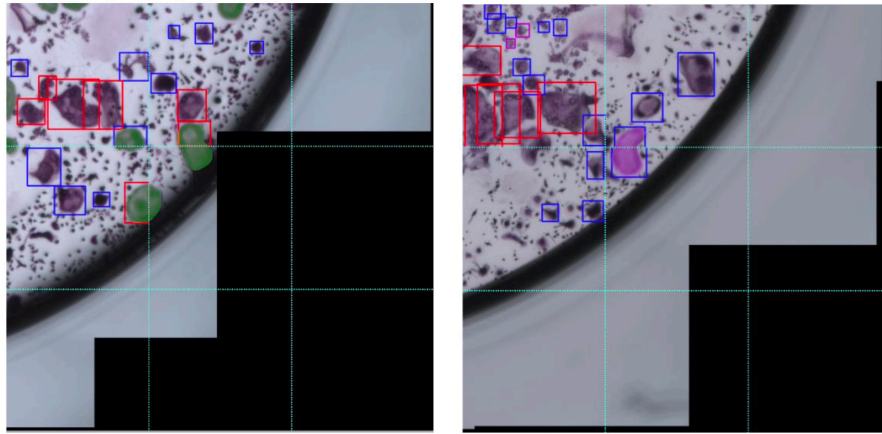


FIGURE 11 | Error in annotation due to overlapping of cells over multiple patches. These two cropped images of cultures are divided into 9 smaller patches. Left: Cells that are split across different patches are marked in opaque green. Right: a single cell split across different patches is marked in opaque violet and is covered by two bounding boxes, one in the upper patch and one in the lower one.

evaluation are the number of cells detected and the area covered by cells. Since these are inherently regression tasks, we report the Pearson correlation of (i) the human inter-annotator agreement for both the number and area covered by TRAP⁺ multinucleated osteoclasts (**Figures 9, 10** upper rows), and (ii) the agreement between the model-predicted and human annotation for these measurements (**Figures 9, 10** bottom rows). In these correlation analyses, the residuals were all normally distributed (as calculated using the D'Agostino & Pearson test in Prism 7.0, $\alpha = 0.05$, $p > 0.4$).

RESULTS

Inter Annotator Agreement

To verify meaningful annotations, we asked three annotators to annotate a test culture divided into 14 regions (the test culture consists of 16 regions, 2 regions did not contain any cells). We then evaluated the inter annotator agreement for the detection and analysis of preosteoclasts, Type 1 and 2 osteoclasts, and ghost cells (**Figures 9, 10** upper rows, **Tables 1, 2**), which is an important measure of the accuracy and reproducibility of the labeling process. Because the most relevant osteoclastic cell type is a TRAP⁺ osteoclast with 3 to 15 nuclei (Type 1), it was important that our model discriminate between preosteoclasts and Type 1 cells as well as between Type 1 and Type 2 osteoclasts. The correlation between the 3 annotators was near-perfect for Type 1 and Type 2 osteoclast counting and area (>0.915), but more moderate coefficients were obtained for the detection of preosteoclasts (from 0.506 to 0.865). The discrepancies in annotation were mainly due to differences in the exhaustiveness of the labeling, which mostly affects the counting of small cells. Since the contribution of the small cells to the estimated area is low, there is still a high level of agreement for area measurements. Notably, the agreement among the human annotators on the

identification of Ghost cells was generally low (0 to 0.917, **Tables 1, 2**).

Model vs. Human Annotation Agreement

Next, to assess the prediction accuracy of our model, we evaluated the agreement between our model and the three human annotators on the same set of images. These images were from a test set that was not used for training purposes. Overall, we found a high level of agreement between the model and human annotators (**Figures 9, 10** bottom row, **Tables 1, 2**). When running our model on the same images used for the inter-annotator correlation, the model agreed with all 3 annotators with correlation coefficients >0.7 for preosteoclasts and Type 1 and 2 osteoclasts. Notably, the correlation among the three human annotators was similar and sometimes even inferior to the model-to-human correlation. In line with the inter-annotator low level of agreement on the identification of Ghost cells, the model's correlation with the human annotators was also low (**Tables 1, 2**). For osteoclast (Type 1 and 2) counting, the correlation between the model and the human annotators ranged from 0.913 to 0.968, and for the area between 0.773 and 0.995 (**Figures 9, 10** bottom row, **Tables 1, 2**). These correlation coefficients are comparable to the inter-annotator coefficients that ranged from 0.948 to 0.960, and from 0.915 to 0.991, respectively (**Figures 9, 10** top row, **Tables 1, 2**).

Error Analysis

In this section we describe the errors that our model makes. The most common issue is with cells that span multiple patches. This problem is introduced due to the resolution at which the model operates and is most frequently evident in bounding boxes that only partially cover the designated cells, as depicted in **Figure 11** [left]. This type of error has a minor effect on the cell counting task, as most cells are covered by a single bounding box. In the area estimation task, this type of error causes partial

coverage of the cells. In rare cases, a single cell is covered by two bounding boxes (See **Figure 11** [right]), which causes double counting of a cell and has no effect on the area estimation task. A visual inspection of the predictions on the test images resulted in 15 erroneous cells from a total of 896 detected cells, which corresponds to 1.5% of the cells in the entire well. The impact of this type of error is therefore negligible.

DISCUSSION AND CONCLUSION

We describe a novel method for rapidly and automatically quantifying osteoclast cultures that uses deep learning methods developed for object detection. Previously developed AI-based approaches for cell detection do not recognize the unique characteristics of osteoclasts and do not differentiate between cell types based on their nuclei number. For the training step, we manually annotated the type and location of each cell in 11 full cultures containing thousands of cells. The novelty of our strategy is that we train an SSD by working at “patch-level” instead of on the full culture and thereby generate more data for the algorithm to train on. The results indicate that our trained model is in good agreement with the human annotators, with a correlation that is similar to inter-annotator correlation. Our model performed especially well for the measurement of Type 1 and Type 2 osteoclast numbers and area. Measurements of preosteoclasts were slightly less satisfactory, although the model agreed with the human annotator to a similar degree and sometimes better than human annotators agreed among themselves. The identification of Ghost cells seems to be particularly problematic with little agreement among the human annotators or between the annotators and the model for these cells.

It should be noted that other experimental settings, e.g., using human cells, and different staining protocols, grayscale images, or camera resolutions, may require dedicated training of the algorithm. In such instances, we suggest that the use of the protocol reported here to generate large numbers of training images will provide a customized model that will perform satisfactorily in each specific setting. Using the same approach, the model could be further improved to detect subclasses of osteoclasts, i.e., 3 to 5 nuclei versus 6 to 10 and 11 to 15 nuclei. In theory, the model could also be trained to recognize and measure fluorescence-stained cultures.

In conclusion, we have developed a satisfactory method for the automation of osteoclast culture analysis that can detect and quantify TRAP-positive, multinucleated osteoclasts. This model discriminates between classical osteoclasts (3 to 15

nuclei) and abnormal “giant” cells (>15 nuclei). This model is therefore a useful labor-saving technique and we suggest that a similar approach may prove beneficial in facilitating other image related analysis tasks.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

EC-K, AB, SE, and ON: algorithm design and development. ZA: experiment design and development. ZA, MK, SBY, and HS: data curation. EC-K, ZA, YG, DN, AG, and YM: data analysis. EC-K, ZA, YG, DN, AG, and YM: writing. YG, DN, AG, and YM: study design and co-direction. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Yandex Machine Learning Initiative at Tel Aviv University, an Israel Science Foundation (ISF) Grant No. 1086/17 to YG and Grant No. 343/17 to DN and by a grant from the Dotan Hemato-oncology Fund, the Cancer Biology Research Center, Tel Aviv University to DN and YG.

ACKNOWLEDGMENTS

This work was carried out in partial fulfillment of the requirements for a Ph.D. degree for ZA and EC-K. The work of AB, SE, and ON was carried out as part of a machine learning workshop, course no. 0368-3300. MK performed an internship with YG and DN in the framework of the “Israel Experience” program and with SBY in the framework of the Sackler Faculty of Medicine’s TAU-MED program.

REFERENCES

- Boyle, W. J., Simonet, W. S., and Lacey, D. L. (2003). Osteoclast differentiation and activation. *Nature* 423, 337–342. doi: 10.1038/nature01658
- Bruzzaniti, A., and Baron, R. (2006). Molecular regulation of osteoclast activity. *Rev. Endocr. Metab. Disord.* 7, 123–139. doi: 10.1007/s11154-006-9009-x
- Ciresan, D., Giusti, A., Gambardella, L., and Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Adv. Neural Inf. Process. Syst.* 25, 2843–2851.
- Clarke, B. (2008). Normal bone anatomy and physiology. *Clin. J. Am. Soc. Nephrol.* 3(Suppl. 3), S131–S139.
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: a retrospective. *Int. J. Comput. Vision* 111, 98–136. doi: 10.1007/s11263-014-0733-5
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1627–1645. doi: 10.1109/tpami.2009.167
- Girshick, R. (2015). “Fast r-cnn,” in *Proceedings of the IEEE International Conference on Computer Vision*, (Washington, DC: IEEE).

- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Columbus, OH: IEEE).
- Hata, K., Kukita, T., Akamine, A., Kukita, A., and Kurisu, K. (1992). Trypsinized osteoclast-like multinucleated cells formed in rat bone marrow cultures efficiently form resorption lacunae on dentine. *Bone* 13, 139–146. doi: 10.1016/8756-3282(92)90003-f
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, NV: IEEE).
- Hiram-Bab, S., Liron, T., Deshet-Unger, N., Mittelman, M., Gassmann, M., Rauner, M., et al. (2015). Erythropoietin directly stimulates osteoclast precursors and induces bone loss. *The FASEB Journal* 29, 1890–1900. doi: 10.1096/fj.14-259085
- Jia, Z., Saxena, A., and Chen, T. (2011). "Robotic object detection: Learning to improve the classifiers using sparse graphs for path planning," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, (Barcelona: IJCAI).
- Karsenty, G., Kronenberg, H. M., and Settembre, C. (2009). Genetic control of bone formation. *Annu. Rev. Cell Dev. Biol.* 25, 629–648. doi: 10.1146/annurev.cellbio.042308.113308
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv [preprint]* Available online at: <http://arxiv.org/pdf/1412.6980v4.pdf> (Accessed March 3, 2015) arXiv:1412.6980,
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.
- Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., et al. (2018). xvnet: objects in context in overhead imagery. *arXiv [preprint]* Available online at: <https://arxiv.org/abs/1802.07856> (Accessed February 22, 2018) arXiv:1802.07856,
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: common objects in context," in *Proceedings of the European Conference on Computer Vision*, (New York, NY: Springer).
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, (New York, NY: Springer).
- Marino, S., Logan, J. G., Mellis, D., and Capulli, M. (2014). Generation and culture of osteoclasts. *BoneKey Rep.* 3:570.
- Menze, M., and Geiger, A. (2015). "Object scene flow for autonomous vehicles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Boston, MA: IEEE).
- Minkin, C. (1982). Bone acid phosphatase: tartrate-resistant acid phosphatase as a marker of osteoclast function. *Calcif. Tissue Int.* 34, 285–290. doi: 10.1007/bf02411252
- Novack, D. V., and Teitelbaum, S. L. (2008). The osteoclast: friend or foe? *Annu. Rev. Pathol. Mech. Dis.* 3, 457–484.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. *arXiv [preprint]* arXiv:1912.01703.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *arXiv [preprint]* Available online at: <https://arxiv.org/abs/1506.01497> (Accessed June 4, 2015) arXiv:1506.01497,
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (New York, NY: Springer), 234–241.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. vision* 115, 211–252. doi: 10.1007/s11263-015-0816-y
- Shorten, C., and Khoshgofaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6, 1–48.
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [preprint]* Available online at: <https://arxiv.org/abs/1409.1556> (Accessed September 4, 2014). arXiv:1409.1556,
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Boston, MA: IEEE).
- Takayanagi, H. (2007). Osteoimmunology: shared mechanisms and crosstalk between the immune and bone systems. *Nat. Rev. Immunol.* 7, 292–304. doi: 10.1038/nri2062
- Takeshita, S., Kaji, K., and Kudo, A. (2000). Identification and characterization of the new osteoclast progenitor with macrophage phenotypes being able to differentiate into mature osteoclasts. *J. Bone Miner. Res.* 15, 1477–1488. doi: 10.1359/jbmr.2000.15.8.1477
- Teitelbaum, S. L. (2000). Bone resorption by osteoclasts. *Science* 289, 1504–1508. doi: 10.1126/science.289.5484.1504
- Teitelbaum, S. L. (2007). Osteoclasts: what do they do and how do they do it? *Am. J. Pathol.* 170, 427–435. doi: 10.2353/ajpath.2007.060834
- Teitelbaum, S. L., and Ross, F. P. (2003). Genetic regulation of osteoclast development and function. *Nat. Rev. Genet.* 4, 638–649. doi: 10.1038/nrg1122
- Wada, K. (2016). *Labelme: Image Polygonal Annotation With Python*. GitHub repository Available online at: <https://github.com/wkentaro/labelme>
- Xue, Y., and Ray, N. (2017). Cell Detection in microscopy images with deep convolutional neural network and compressed sensing. *arXiv [preprint]* Available online at: <https://arxiv.org/abs/1708.03307> (Accessed February 21, 2018). arXiv:1708.03307,
- Zaidi, M. (2007). Skeletal remodeling in health and disease. *Nat. Med.* 13, 791–801. doi: 10.1038/nm1593

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cohen-Karlik, Awida, Bergman, Eshed, Nestor, Kadashev, Yosef, Saed, Mansour, Globerson, Neumann and Gabet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.