Assessing the Performance of a Machine Learning Algorithm in
Identifying Bubbles in Dust Emission

Duo Xu (许铎) and Stella S. R. Offner
1 Department of Astronomy, University of Massachusetts, Amherst, MA 01002, USA
2 Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA; xuduo117@utexas.edu, soffner@astro.as.utexas.edu

Abstract
Stellar feedback created by radiation and winds from massive stars plays a significant role in both physical and chemical evolution of molecular clouds. This energy and momentum leaves an identifiable signature (“bubbles”) that affects the dynamics and structure of the cloud. Most bubble searches are performed “by eye,” which is usually time-consuming, subjective, and difficult to calibrate. Automatic classifications based on machine learning make it possible to perform systematic, quantifiable, and repeatable searches for bubbles. We employ a previously developed machine learning algorithm, Brut, and quantitatively evaluate its performance in identifying bubbles using synthetic dust observations. We adopt magnetohydrodynamics simulations, which model stellar winds launching within turbulent molecular clouds, as an input to generate synthetic images. We use a publicly available three-dimensional dust continuum Monte Carlo radiative transfer code, HYPERION, to generate synthetic images of bubbles in three Spitzer bands (4.5, 8, and 24 μm). We designate half of our synthetic bubbles as a training set, which we use to train Brut along with citizen-scientist data from the Milky Way Project (MWP). We then assess Brut’s accuracy using the remaining synthetic observations. We find that Brut’s performance after retraining increases significantly, and it is able to identify yellow bubbles, which are likely associated with B-type stars. Brut continues to perform well on previously identified high-score bubbles, and over 10% of the MWP bubbles are reclassified as high-confidence bubbles, which were previously marginal or ambiguous detections in the MWP data. We also investigate the influence of the size of the training set, dust model, evolutionary stage, and background noise on bubble identification.

Key words: ISM: bubbles – ISM: clouds – methods: data analysis – stars: formation

1. Introduction

During the process of star formation, stellar feedback plays a significant role in both the physical and chemical evolution of molecular clouds (Hollenbach & Tielens 1999; Frank et al. 2014). One of the most important feedback mechanisms is mass loss (Lada 1985). There are two typical manifestations of stellar winds: protostellar outflows, which are often highly collimated, and radiatively driven winds from main-sequence stars, which are more isotropic (Churchwell et al. 2006; Arce et al. 2010, 2011; Li et al. 2015). Both types of stellar winds inject momentum and energy into the environment, and thereby affect the dynamics and structure of the parent molecular cloud.

Recent observational studies have shown that the momentum and energy injected by stellar winds are one or more orders of magnitude larger than those of outflows owing to their larger volume and longer lifetime (Arce et al. 2011; Li et al. 2015). Arce et al. (2011) found that the energy injection rate from these stellar winds is comparable to the turbulent dissipation rate in the Perseus molecular cloud, which means that in the current epoch, stellar feedback is sufficient to maintain the observed turbulence in Perseus. A similar conclusion was also reached by Li et al. (2015) in the Taurus molecular cloud. It is notable that both regions are low-mass star-forming regions, and that high-mass stars, which generally dominate feedback energetics, are absent.

Simulations confirm the significant kinematic impact due to stellar feedback on the global star formation process. Winds can replenish energy dissipated by turbulence and also trigger star formation by compressing the cloud (Matzner 2002; Dale et al. 2005, 2013, 2014; Nakamura & Li 2007; Dale & Bonnell 2008; Wang et al. 2010). Winds can also gradually ablate the molecular material from forming stellar clusters (Rogers & Pittard 2013). Offner & Arce (2015) quantified the stellar wind mass-loss rates for individual stars, which they found must be greater than $10^{-7} M_{\odot}$ yr$^{-1}$ to be consistent with observations. Additionally, ionizing radiation feedback from O stars also influences the morphology of clouds and the formation of stars (Dale et al. 2005, 2013, 2014; Geen et al. 2015; Kim et al. 2016).

Despite many observational and theoretical studies, the importance and impact of feedback on molecular clouds remain debated. This is because wind signatures are difficult to identify and quantify. Most bubble searches are done “by eye” (Churchwell et al. 2006; Arce et al. 2011; Li et al. 2015). For example, over 35,000 citizen scientists participated in the Milky Way Project (MWP, Simpson et al. 2012) in order to identify bubbles in Spitzer images. This approach is time-consuming, subjective, and difficult to calibrate (Beaumont et al. 2014). Analyzing the completeness of visually identified bubbles, which has a significant effect on the estimation of the injected momentum and energy, remains a great challenge. However, automatic classifications driven by machine learning approaches enable systematic, quantifiable, and repeatable searches to identify bubbles (Beaumont et al. 2011, 2014).

One of the most popular types of machine learning algorithms in astronomical classification is “Random Forests” (e.g., Carliles et al. 2010; Beaumont et al. 2014; Masci et al. 2014), which are based on decision trees. A decision tree is a data structure that classifies feature vectors by computing a series of constraints, and propagating vectors down the tree based on whether these constraints are satisfied. Compared to
other machine learning approaches, the Random Forests approach does well in classifying problems that have a large number of feature dimensions (Breiman 2001). Beaumont et al. (2014) developed an algorithm Brut based on Random Forests and applied it to classifying bubbles in the Milky Way. For each bubble, they defined a “score,” which is related to the probability that a given structure is a bubble. After conducting a blind search in the Milky Way, they found a substantial population of low-score bubble candidates not in the MWP catalog produced by citizen scientists. In other words, citizen scientists are likely to miss a significant number of bubbles, but machine leaning can compensate for some of this incompleteness.

Increasingly rich and detailed data on the local interstellar medium (ISM) and star-forming regions are available, such as the Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE, Benjamin et al. 2003), the Hi-GAL (Herschel infrared Galactic Plane, Molinari et al. 2010) Survey, and the GALFA-Hi (The Galactic Arecibo L-band Feed Array HI, Peek et al. 2011) Survey. Parsing these extensive data visually is prohibitively time-consuming but is possible with the aid of machine learning algorithms.

There are two main types of machine learning algorithms: unsupervised learning and supervised learning. Unsupervised learning algorithms make their own criteria to discover structure in the data. An algorithm that learns from a training data set and makes decisions based on the input “knowledge” is called supervised learning. Supervised learning iteratively makes predictions on the training data and is corrected by the input training data set. Consequently, the training data set plays a significant role in the accuracy with which the task is performed.

One fundamental problem with visual identification is that bubbles identified “by eye” are not objective and can be incorrect, such that machine learning approaches trained using flawed visual data will in turn produce defective identifications. In addition, there is no independent, quantitative assessment for completeness or any clear metric to determine how well bubbles are actually identified. One solution is to use realistic simulations, where feedback properties are known and well defined. Such simulations can evaluate the accuracy of the training data and in turn supplement the original training data set.

In this paper, we assess the performance of Brut in identifying bubbles using synthetic observations. We produce synthetic dust observations of bubbles in simulations. We use these as a supplemental training set to retrain Brut and test the performance of retrained Brut in classifying both synthetic bubbles and observed bubbles. We describe the method we use to construct synthetic observations and the details of the machine learning algorithm in Section 2. We compare and discuss several synthetic observation models in Section 3. In Section 4, we present the performance of retrained Brut in classifying both synthetic bubbles and observed bubbles. We summarize our results and conclusions in Section 5.

### 2. Methods

#### 2.1. Hydrodynamic Simulations

We adopt the magnetohydrodynamics simulations from Offner & Arce (2015), which aim to model winds from intermediate-mass stars and explore their impact on cloud morphology and turbulence. The simulations model a piece of a molecular cloud with length $L = 5$ pc, mass $M = 3762 M_\odot$, and periodic boundary conditions. The initial cloud temperature is $T = 10$ K. The initial density and velocity conditions are set through driving the gas without gravity by adding random large-scale perturbations to the velocity field. These simulations share the same Alfvén Mach number 2.3 but their magnetic field distributions are spatially different at the initial time. Their Fourier spectral slopes of velocity and density are comparable to $S(k) \propto k^{-1.7}$ and $S(k) \propto k^{-1.3}$, respectively. The turbulence is initially external driving but ceases when the stellar sources are inserted and the feedback begins. Table 2 lists the parameters of these models. More details about the simulations can be found in Offner & Arce (2015).

We adopt outputs from the strong wind run in which five stellar sources with different mass-loss rates are placed randomly. The number density of sources is similar to that in Perseus. These sources are all B-type stars with mass-loss rates ranging from $2.6 \times 10^{-5}$ to $1.8 \times 10^{-5} M_\odot$ yr$^{-1}$. Table 1 lists the physical parameters of each of the five stellar sources. In this work, we explored outputs with different stages of evolution and different realizations of turbulence.

#### 2.2. Hyperion

We use the publicly available three-dimensional dust continuum Monte Carlo radiative transfer code HYPERION (Robitaille 2011) to generate synthetic observations of the simulations described in Section 2.1. We adopt the gas density and temperature distributions from the outputs listed in Table 2 and the stellar properties from Table 1 as inputs. HYPERION assumes that a star radiates as a blackbody.

Assumptions about the dust properties strongly influence the resulting emission. A variety of models for ISM dust have been proposed in the literature (e.g., Kim et al. 1994; Draine 2003; Koeperl et al. 2017), and we explore four different models in this work. Following Koeperl et al. (2017), we combine three different dust grain models with 80.63% big grains ($>200$ Å),

---

**Table 1**

| ID | $M (M_\odot)$ | $L (10^3 L_\odot)$ | $T (10^4$ K) | $M (10^{-7} M_\odot$ yr$^{-1}$) |
|----|---------------|---------------------|--------------|---------------------------------|
| 1  | 3.8           | 0.19                | 2.3          | 0.35                            |
| 2  | 10.4          | 6.3                 | 3.8          | 9.1                             |
| 3  | 12.2          | 10.3                | 3.6          | 17.7                            |
| 4  | 13.1          | 12.8                | 3.1          | 12.4                            |
| 5  | 12.4          | 10.8                | 2.6          | 2.5                             |

**Table 2**

| Model | $t_1$ ($t_{zone}$) | $t_{min}$ (Myr) |
|-------|-------------------|-----------------|
| T1_t1 | 1.6               | 0.1             |
| T2_t1 | 2.0               | 0.1             |
| T2_t0 | 2.0               | 0.05            |

**Notes.**

a) Model name, the initial start time in crossing times, and the evolutionary time. All models have $L = 5$ pc, $M = 3762 M_\odot$, $T = 10$ K, and initial $B = 13.5 \mu$G.

b) Output corresponding to the model “W1_T1” with an evolutionary time of 0.1 Myr in Offner & Arce (2015).

c) Output corresponding to the model “W1_T2” in Offner & Arce (2015).
13.51% smaller dust species, called very small grains (vsg, 20–200 Å), and 5.86% polycyclic aromatic hydrocarbon (PAH) molecules, called ultrasmall grains (usg, <20 Å). We label this dust model “K16” in the following discussion. We assume a moderate gas-to-dust ratio of 100 (Savage & Mathis 1979) and adopt a regular Cartesian grid with young stars embedded within. We calculate the emission for 20 different angular views and convolve the spectra with the Spitzer transmission curve (Quijada et al. 2004; MIPS Instrument & MIPS Instrument Support Teams 2011) to generate synthetic images in three Spitzer bands (4.5, 8, 24 μm) at doi:10.7910/DVN/OSMNDG. Figure 1 shows synthetic bubble images of the five sources with 20 different viewing angles.

In addition to the K16 dust model above, we adopt three other commonly used dust models to produce synthetic observations:

(1) the “kmh” dust model (Kim et al. 1994), which consists of astronomical silicates, graphite, and carbon with full scattering properties;
(2) the “Draine” dust model (Draine 2003), which is mainly carbonaceous-silicate grains in the Milky Way;
(3) the “IPS” dust model (Semenov et al. 2003), which represents “iron-poor” silicate dust.

Figure 2 shows synthetic images adopting the kmh dust model. The synthetic observations adopting the Draine and IPS dust models are similar to these, so we only include images with the kmh model.

The spectral energy distributions (SEDs) of different dust models show distinct differences, especially at 8 μm where PAH emission dominates. We extract the observed spectra of the main molecular cloud of Ophiuchus, LDN 1688 (Rawlings et al. 2013), and compare the SEDs of the different dust models as shown in Figure 3. The K16 dust model appears to be more realistic since it includes PAH emission while the other models lack PAH emission around 8 μm. Since the SEDs of the kmh model, Draine model, and IPS model have a similar intensity at 4.5, 8, and 24 μm, the Draine and IPS three-color synthetic images look similar to the kmh model shown in Figure 2. The
The morphology of the bubbles changes with time as the winds expand into the cloud and interact with the surrounding gas. The bubbles at earlier evolutionary stages are more compact than those at later stages, which have undergone additional expansion driven by the stellar wind. Figure 6 shows younger bubbles (“T2_t0” listed in Table 2). The bubbles at the earlier time appear brighter in the center, owing to their compact and concentrated structure.

3.3. Turbulent Realization

We also analyze a simulation with different initial turbulence. The synthetic observation process remains the same as described above, where we crop the HYPERION input data cube and use the K16 dust model. Figure 7 shows the synthetic images with different initial turbulence (“T1_t1” listed in Table 2). “T1_t1” and “T2_t1” have the same initial mean magnetic field, ratio of thermal to magnetic pressure, mean density, and stellar properties, but the shapes of the bubbles are distinctly different owing to the different density distribution of the cloud material. Since the turbulent structure of real molecular clouds is varied, we adopt different initial turbulence to explore the diversity of bubble morphology and enrich our training data set.

3.4. Noise

The synthetic images are smooth, unlike real observational images, which have fluctuations produced by noise. It is important that the training data be as close as possible to the observational data to reduce bias in detection caused by differences. To make the synthetic images more realistic, we identify patches of GLIMPSE data that are removed from the Galactic plane and have low signal-to-noise ratio (S/N). We add these “stamps” to the synthetic images using the same S/N as the GLIMPSE data, where S/N ~ 8. Figure 8 shows the synthetic bubble images with noise.

Table 3

| Random Forest Zone | Training Zone (°) | Test Zone (°) |
|--------------------|------------------|---------------|
| r1                 | 3n + 0°5 ≤ l < 3n + 1°5 | 3n + 1°5 < l < 3n + 3°5 |
| r2                 | 3n + 1°5 ≤ l < 3n + 2°5 | 3n – 0°5 ≤ l < 3n + 1°5 |
| r3                 | 3n – 0°5 ≤ l < 3n + 0°5 | 3n + 0°5 < l < 3n + 2°5 |

Notes.
- a The training zones are interleaved across all longitudes.
- b n is an integer ranging from 0 to 119.
- c When n is 119, the test zone is 358°5 < l < 0°5(360°5).
- d When n is 0, the training zone is 359°5(–0°5) ≤ l ≤ 0°5.

...
Figure 4. Three-color synthetic images adopting the K16 dust model where the HYPERION input is cropped to 2.2 pc.

Figure 5. Three-color synthetic images adopting the K16 dust model where the HYPERION input is cropped to 3 pc.

Figure 6. Three-color synthetic images at an earlier evolutionary stage with 0.05 Myr (“T2_t0” listed in Table 2) where the HYPERION input is cropped to 3 pc.

Figure 7. Three-color synthetic images with different turbulence where the HYPERION input is cropped to 3 pc.
4. Results

4.1. Retraining Brut with Synthetic Observation

We divide all the synthetic images into two equal parts. One half acts as a training data set, which we use to supplement the original MWP bubble set. The other serves as a test set, which allows us to assess the performance of the retrained algorithm. We summarize all the synthetic images we use in the training and testing sets in Table 4.

We analyze the performance of the three Random Forests before and after supplementing them with the new training data. First, we retrain Brut using the synthetic images without noise (IDs 1–7 in Table 4). Figure 9 shows the performance with the original training and with the algorithm retrained on noiseless synthetic images on the test bubbles. Table 5 briefly describes the meaning of the labels in Figure 9. The scores returned after retraining on the noiseless data are significantly higher than those given by the original training. After retraining, the feature vector more accurately represents the synthetic bubbles and Brut does a better job of identifying them.

We then augment the training set by adding the bubbles with IDs 7–14 in Table 4, so that the new training set consists of half of the bubbles with and without noise. Figure 10 shows the performance with the original training and the algorithm retrained on synthetic images with and without noise on the second half of the noisy data. The scores returned by the retrained algorithm are significantly higher than those given with the original training. Compared with the scores retrained using noiseless data in Figure 9, the scores given by the retrained algorithm including some noisy images are more concentrated. This is likely because delicate bubble structure is reduced, i.e., there is less variation in bubble appearance since the noise hides small-scale substructure.

We next explore the impact of the size and composition of the training set on the performance of the retrained algorithm. We retrain the algorithm with only synthetic images and also with a set containing half the number of MWP bubbles and all the synthetic images. Figure 11 shows the performance of the algorithm trained with only synthetic images and the algorithm trained with fewer MWP images + synthetic images on the noisy data. Compared with the scores returned when training with all the MWP data and synthetic images in Figure 10, the scores returned by different random forests are similar but more concentrated. This is likely caused by the larger fraction of synthetic bubbles, which are similar to the test set, in the training set. The synthetic images are responsible for the better performance of the retrained algorithm on the test set of synthetic images.

The increased scores after retraining suggest that the original training data set is incomplete, especially lacking bubbles driven by intermediate or low-mass stars. We further examine

---

**Figure 8.** Three-color synthetic images with noise adopting the K16 dust model where the HYPERION input is cropped to 3 pc.

**Table 4**

Parameters of the Synthetic Images

| ID | Labela | Turbularityb | Evolutionary Stage (Myr) | Image Size (pc) | Crop | Noise |
|----|--------|--------------|--------------------------|----------------|------|-------|
| 1  | T1_t1_c2 | T1           | 0.1                      | 2.2            | ✓    | X     |
| 2  | T1_t1_c3 | T1           | 0.1                      | 3              | ✓    | X     |
| 3  | T2_t1_2  | T2           | 0.1                      | 2.2            | X    | X     |
| 4  | T2_t0_c2 | T2           | 0.05                     | 2.2            | ✓    | X     |
| 5  | T2_t0_c3 | T2           | 0.05                     | 3              | ✓    | X     |
| 6  | T2_t1_c2 | T2           | 0.1                      | 2.2            | ✓    | X     |
| 7  | T2_t1_c3 | T2           | 0.1                      | 3              | ✓    | X     |
| 8  | T1_t1_c2n| T1           | 0.1                      | 2.2            | ✓    | ✓     |
| 9  | T1_t1_c3n| T1           | 0.1                      | 3              | ✓    | ✓     |
| 10 | T2_t1_c2n| T2           | 0.1                      | 2.2            | ✓    | ✓     |
| 11 | T2_t0_c2n| T2           | 0.05                     | 2.2            | ✓    | ✓     |
| 12 | T2_t0_c3n| T2           | 0.05                     | 3              | ✓    | ✓     |
| 13 | T2_t1_c2n| T2           | 0.1                      | 2.2            | ✓    | ✓     |
| 14 | T2_t1_c3n| T2           | 0.1                      | 3              | ✓    | ✓     |

**Notes.**

a A label with “n” indicates a synthetic image with noise.

b Turbulent distributions listed in Table 2.
the performance of retrained Brut on observational data in Sections 4.2 and 4.3.

4.2. Retesting Brut on the MWP Data

We adopt all 3716 large bubbles found by the citizen scientists in Simpson et al. (2012) as a test set to assess the performance of Brut after retraining. We ignore the objects contained in the “small bubble” catalog, which are mainly green knots, dark nebulae, star clusters, galaxies, or fuzzy red objects. We compare the performance for the original training, the retrained algorithm (using both noisy and noiseless synthetic images), and the algorithm trained with fewer MWP images + synthetic images in classifying MWP bubbles, as shown in Figure 12. The scores returned by the retrained algorithm are significantly higher than those returned by Brut without additional training. When we retrain the algorithm with only synthetic images, the scores under 0.55 show a dramatic improvement. After investigating the bubble images with high and low scores, we find that the algorithm trained with only synthetic images improves the scores of ambiguous bubbles with low S/N and reduces the scores of red bubbles with high S/N.

To explain the performance of the algorithm retrained on several different training sets, we characterize the bubble properties that compose each training set as shown in Figure 13. The “Normalized S/N” quantifies the contrast and...
$S/N$ of the image. We define it as

$$\text{Normalized } S/N = C (I_{95} - I_{30})(I_{95} - I_{00}) f_{8\sigma},$$

where $(I_{95} - I_{30})$ is the difference between the top 5% and the bottom 30% of values, $(I_{95} - I_{00})$ is the difference between the top 5% values and the median value, $f_{8\sigma}$ is the fraction of bright pixels ($\geq 8\sigma$), and $C$ is a constant to normalize the values to unity. In most of the high-$S/N$ bubble images, the bubble rim structures occupy the top 5% of the image values, and the noise occupies the bottom 30%. The average of the diffuse emission is well represented by the median image value. We use this product to indicate the contrast of the image. In a random noisy image, the normalized $S/N$ is close to 0. The $x$-axis in Figure 13 indicates the “Yellow Index,” which describes the color of the bubble. We define it as the ratio between the number of yellow pixels and the number of red pixels. Although the original training set spans a wide range of color and $S/N$, it is concentrated in the red domain. The large representation of red bubbles in the training set means that Brut will more easily identify red bubbles than yellow bubbles. In contrast, the synthetic bubbles are located in the yellow part of the parameter space. The MWP bubbles are mostly low-$S/N$ red bubbles, with some low-$S/N$ yellow bubbles and high-$S/N$ red bubbles, but there are very few high-$S/N$ yellow bubbles. Consequently, the algorithm trained with only synthetic images mainly captures bubbles with low $S/N$. This explains why a
training set with only synthetic images improves the scores of ambiguous bubbles with low $S/N$ and reduces the scores of bubbles that are red and have high $S/N$.

Figure 12 also shows the result when we randomly remove half of the bubbles in the original MWP training set. The score distribution returned by the algorithm trained with fewer MWP images + synthetic images is surprisingly different from that returned by the algorithm trained with only synthetic images. When including half of the original MWP bubbles in the training set, the performance of the algorithm dramatically decreases. The original MWP training bubbles are mostly red, while the synthetic images nearly all contain yellow bubbles. Consequently, these sets inhabit two different color domains. The reduction of red bubbles in the training set lowers the scores of these types of bubbles in the test set. When including all the original and synthetic images in the training set, the performance of the retrained algorithm significantly and steadily improves. Consequently, this demonstrates that the composition and size of the training set significantly impacts the performance of the algorithm.

Following our comparison of the performance of the algorithm after retraining with several different training sets, we adopt the training that includes all the original MWP bubbles and synthetic images. This training set significantly improves the scores of most bubbles with little change in the number of high-score objects. Although the algorithm trained with only synthetic images improves the scores of a large number of bubbles, it no longer returns any high-score bubbles, which were previously assigned to images with red bubbles.

The MWP characterizes the consensus among users that an image contains a bubble in terms of the “hit rate,” which is the fraction of citizen scientists who identified a bubble in the image. They define hit rates above 0.1 as being high-confidence bubble candidates.

We further compare the scores given by Brut with the original training, the scores after it is retrained, and the MWP hit rate, as shown in Figure 14. The average Brut score in each bin with the original training and that after retraining both show a clear trend with the hit rate. The error bars indicate the standard deviation of the scores and hit rate in each bin. The higher the hit rate, the higher the score Brut returns, which is consistent with our expectations. In other words, the retrained algorithm preserves the hit-rate distribution, where bubbles with low hit rates continue to have low scores.

Moreover, over 10% of the MWP bubbles, which were previously marginal or ambiguous detections, are reclassified as high-confidence bubbles after retraining. Their average Brut score increases from −0.07 to 0.39. About 2% of the previously identified MWP bubbles are no longer classified as high-confidence bubbles, and their average Brut score drops from 0.31 to 0.06.

Figure 15 shows 100 bubbles whose score significantly increases after retraining. Most of these bubbles are yellow, indicating that the 8 and 24 $\mu$m emissions are similar. These yellow bubbles are likely ultracompact and compact H II regions or analogous regions for less massive B-type stars (Kerton et al. 2015). The performance of the retrained algorithm is consistent with our training set, in which bubbles are created by the stellar winds of B-type stars. For stars of this
type, the amount of ionizing radiation is small, so the bubbles are predominantly cleared by the wind (or earlier protostellar outflows) and then illuminated by the stellar radiation field. Figure 16 shows nine bubbles, which were previously identified MWP bubbles but are no longer classified as high-confidence bubbles after retraining. These bubbles are very red and thus quite distinct from our yellow bubbles, and their morphology does not show a distinct shell rim. Consequently, since we supplemented the training set with synthetic yellow bubbles, the decline in these bubbles’ Brut scores is unsurprising.

In summary, the performance of the retrained algorithm in classifying yellow bubbles increases significantly when synthetic observations are added to the training set.

4.3. Application: Bubbles in the Perseus Molecular Cloud

Perseus is located in the larger Taurus–Auriga–Perseus dark cloud complex at a distance of 250 ± 50 pc and spans a total area of about 70 pc² (Enoch et al. 2006; Evans et al. 2009). With a mass of 10^4 M_☉, the Perseus cloud is often considered to be an intermediate case between low-mass star-forming regions such as Taurus and turbulent, high-mass regions such as Orion.
Ladd et al. (1994), making it an ideal location to study low- and intermediate-mass star formation. The feedback of young stars makes Perseus a “bubbly” cloud (Arce et al. 2011).

Arce et al. (2011) identified 12 bubbles using CO spectral data. We extract the Spitzer image of Perseus in 4.5, 8, and 24 μm bands (R. Gutermuth 2017, private communication) and apply Brut to these data. Figure 17 shows four examples of bubbles in the Perseus molecular cloud. These bubbles are associated with shells CPS6, CPS8, CPS10, and CPS11 in the CO data, which were visually identified by Arce et al. (2011). Table 6 lists the physical properties of these bubbles. All these bubbles are probably driven by relatively low- or intermediate-mass young stars such as B-type or F-type stars. Figure 17 shows these bubbles and their associated Brut scores before and after retraining. These four cases show a significant improvement in score, mostly from a negative score (non-bubble) to a positive score (likely bubble).

CPS6 and CPS8 are similar to the synthetic bubbles and the MWP yellow bubbles. They are the best examples of the good performance produced by retraining Brut. CPS11 is a partial bubble, which is probably why its score is still <0.2. Table 6 shows that CPS10 is driven by a B5V star, but there is no distinct evidence of the existence of the star in the infrared images. However, the star is bright and clearly visible in the optical data.

The emission from dust exceeds that from the star, so the B5V star becomes invisible when embedded in the cloud. Although CPS10 is not a yellow bubble, it is nonetheless consistent with the bubble model in Figure 1 in Beaumont et al. (2014), where a green shell structure is produced by PAH emission and the red interior is dominated by hot dust. The bubble score is low, which is likely due to the contamination by other emission in the upper right corner.

These results indicate that the retrained algorithm can perform well for molecular cloud data not included in the MWP. The synthetic observations are able to improve Brut’s performance in classifying bubbles produced by relatively low- or intermediate-mass young stars such as B-type stars.

5. Conclusions

We adopt magnetohydrodynamics simulations of stellar winds interacting with a molecular cloud and post-process them using a three-dimensional dust continuum Monte Carlo radiative transfer code. We generate synthetic observations of bubbles in the Spitzer bands (4.5, 8, and 24 μm). We employ a previously developed machine learning algorithm, Brut, and quantitatively evaluate its performance in identifying bubbles using synthetic dust observations. Our main findings are the following.

1. Synthetic observations in combination with visually identified sources can be used to significantly improve machine learning classification.
2. After retraining with synthetic images, Brut better identifies yellow bubbles, which are likely associated with H II regions for less massive B-type stars or cavities evacuated by stellar winds.
3. The completeness of the training set significantly impacts the performance of the algorithm. We suggest that the number of yellow bubbles in the current MWP bubble catalog is incomplete, and we expect that a random search of the full GLIMPSE data set with Brut would return many more yellow bubble candidates.
4. Some of the bubbles with improved scores are associated with sources of lower confidence in the MWP. These would likely be identified as bubbles by an expert, and thus the simulations provide an efficient means to enhance training sets for machine learning.
5. Turbulent structures greatly affect the morphology of bubbles, yielding a variety of bubble shapes. Different evolutionary stages and different cropped image sizes further enhance the bubbles contained in the training set. Adding noise similar to that in the GLIMPSE data makes the synthetic observations more realistic. In combination, these modifications create a more complete training set to improve the machine learning classification.

6. The retrained algorithm performs well in classifying bubbles associated with more embedded sources located in Perseus. Thus, retraining with synthetic observations expands the parameter space of the training set beyond the less embedded and more distant regions with massive stars covered by the MWP.

We thank Rob Gutermuth and Chris Beaumont for helpful comments and contributions. We also thank the anonymous referee for comments that improved this manuscript. S.O. acknowledges support from NSF AAG AST-1510021 and NASA ATP NNX15AT05G.

References

Arce, H. G., Borkin, M. A., Goodman, A. A., Pineda, J. E., & Beaumont, C. N. 2011, ApJ, 742, 105
Arce, H. G., Borkin, M. A., Goodman, A. A., Pineda, J. E., & Halle, M. W. 2010, ApJ, 715, 1170
Beaumont, C. N., Goodman, A. A., Kendrew, S., Williams, J. P., & Simpson, R. 2014, ApJS, 214, 3
Beaumont, C. N., Williams, J. P., & Goodman, A. A. 2011, ApJ, 741, 14
Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, PASP, 115, 953
Breiman, L. 2001, Mach. Learn., 45, 5
Carliles, S., Budavári, T., Heinis, S., Pribe, C., & Szalay, A. S. 2010, ApJ, 712, 511
Churchwell, E., Povich, M. S., Allen, D., et al. 2006, ApJ, 649, 759
Dale, J. E., & Bonnell, I. A. 2008, MNRAS, 391, 2
Dale, J. E., Bonnell, I. A., Clarke, C. J., & Bate, M. R. 2005, MNRAS, 358, 291

Dale, J. E., Ngoumou, J., Ercolano, B., & Bonnell, I. A. 2013, MNRAS, 436, 3430
Dale, J. E., Ngoumou, J., Ercolano, B., & Bonnell, I. A. 2014, MNRAS, 442, 694
Draine, B. T. 2003, ARA&A, 41, 241
Enoch, M. L., Young, K. E., Glenn, J., et al. 2006, ApJ, 638, 293
Evans, N. J., II, Dunham, M. M., Jørgensen, J. K., et al. 2009, ApJS, 181, 321
Frank, A., Ray, T. P., Cabrit, S., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 451
Gegen, S., Rosdahl, J., Blaizot, J., Devriendt, J., & Slyz, A. 2015, MNRAS, 448, 3248
Hollenbach, D. J., & Tielens, A. G. G. M. 1999, RevMP, 71, 173
Kerton, C. R., Wolf-Chase, G., Arvidsson, K., Lintott, C. J., & Simpson, R. J. 2015, ApJ, 799, 153
Kim, J.-G., Kim, W.-T., & Ostriker, E. C. 2016, ApJ, 819, 137
Kim, S.-H., Martin, P. G., & Hendry, P. D. 1994, ApJ, 422, 164
Koepernik, C. M., Robitaille, T. P., Dale, J. E., & Biscani, F. 2017, ApJS, 233, 1
Lada, C. J. 1985, ARA&A, 23, 267
Ladd, E. F., Myers, P. C., & Goodman, A. A. 1994, ApJ, 433, 117
Li, H., Li, D., Qian, L., et al. 2015, ApJ, 812, 20
Masci, F. J., Hoffman, D. L., Grillmair, C. J., & Cutri, R. M. 2014, AJ, 148, 21
Matzner, C. D. 2002, ApJ, 566, 302
MIPS Instrument & MIPS Instrument Support Teams 2011, MIPS Instrument Handbook, Version 3 (Pasadena, CA: SSC), https://irsa.ipac.caltech.edu/data/SPITZER/docs/mips/mipsinstrumenthandbook/MIPS_Instrument_Handbook.pdf
Molnari, S., Swinyard, B., Bally, J., et al. 2010, PASP, 122, 314
Mora, A., Merin, B., Solano, E., et al. 2001, A&A, 378, 116
Nakamura, F., & Li, Z.-Y. 2007, ApJ, 662, 395
Offner, S. S. R., & Arce, H. G. 2015, ApJ, 811, 146
Peek, J. E. G., Heiles, C., Douglas, K. A., et al. 2011, ApJS, 194, 20
Preibisch, T., & Zinnecker, H. 2001, ApJ, 559, 137
Quijada, M. A., Marx, C. T., Arendt, R. G., & Moseley, S. H. 2004, Proc. SPIE, 5487, 244
Rawlings, M. G., Juvela, M., Lehtinen, K., Mattila, K., & Lemke, D. 2013, MNRAS, 428, 2617
Rebull, L. M., Stapelfeldt, K. R., Evans, N. J., II, et al. 2007, ApJS, 171, 447
Robitaille, T. P. 2011, A&A, 536, A79
Rogers, H., & Pittard, J. M. 2013, MNRAS, 431, 1337
Semenov, D., Henning, T., Hollenbach, D., & Schleicher, D. 2003, A&A, 410, 611
Simpson, R. J., Povich, M. S., Kendrew, S., et al. 2012, MNRAS, 424, 2442
Wang, P., Li, Z.-Y., Abel, T., & Nakamura, F. 2010, ApJ, 709, 27