1

Computer Vision in Chemical Reaction Monitoring and Analysis

*Marc Reid**

UKRI Future Leaders Fellow, Department of Pure & Applied Chemistry, University of Strathclyde, Glasgow, UK

*Corresponding author: marc.reid.100@strath.ac.uk

1.1 Introduction

In February 2015, the Internet was consumed by an argument about fashion. A particular photograph of a particular dress was showing up wherever you might have scrolled on social media. Hanging on the rail of a shop, captured under a stream of daylight, this dress appeared, to many, to be white and gold. Yet, as if to represent the entire other half of the online community, others saw the same photo of the same dress as being blue and black. To this day, you can share that same photograph with unsuspecting live audiences, and by a show of hands, will recreate the eerie bipartisan opinion regarding the colors of the dress. But who was correct? Is the dress white and gold, blue and black, or some other combination in between?

The dress is, in fact, blue and black. While the consensus on the root cause of the debate is not clear [1], psychological variation in human perception of colors, and our predicted corrections of those colors, likely plays a role. An illustration of the phenomenon for the reader of this book lies on the right side of Figure 1.1. For the four regions of interest (ROI) highlighted, Table 1.1 shares the average colors expressed in terms of mathematical color models we will discuss in the next section. For now, note that, for regions A and D, the numerical expressions of their colors

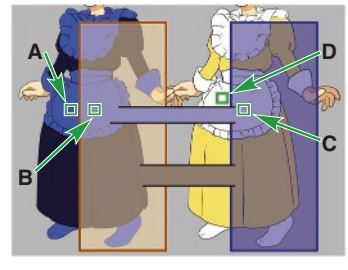

Figure 1.1 Left: the original dress image that sparked the Internet debate. Right: cartoon variant of the dress, shows changes in likely color perception under different light sources. The four individual square ROI (left to right on the cartoon) mark areas A–D described in Table 1.1.

Artificial Intelligence in Catalysis: Experimental and Computational Methodologies, First Edition.
Edited by Valentine P. Ananikov and Mikhail V. Polynski.

Table 1.1 Average RGB, HSV, and CIE–L*a*b* pixel values for four 120 × 120 square ROI on the cartoon apron, under each light source shown in Figure 1.1.

ROI[a]	Average RGB	Average HSV	Average L*a*b*
A	[38, 59, 222]	[53, 211, 222]	[54, 43, −63]
B	[147, 140, 187]	[68, 64, 187]	[79, 17, −10]
C	[147, 140, 187]	[68, 64, 187]	[79, 17, −10]
D	[255, 254, 255]	[300, 1,255]	[99, 8, 5]

a. All ROI represent 120 × 120 pixels.

are not identical for this pair of regions. However, for regions B and C, each showing the same apron under a warm and cold shade of light, show identical numerical expressions of their color, despite what your brain may tell you is different when judging these colors by eye.

While the story of the dress is not meant to set this chapter up as one about fashion history, it instead serves to demonstrate the subjectivity of our human perception of all things color. Furthermore, we see that the quantification of color, exemplified in Table 1.1, begins to reveal the more objective expression of color that lies "under the hood" of how it is captured by a camera. This subjective-to-objective shift is vital to enable a trustworthy and exploitable analysis of color-based phenomena.

Instead of formalwear, we might instead be concerned about the rate at which a palladium catalyst degrades into its notorious nightshade, palladium black. Equally, we could be making a note of the curious colors emerging from a settling bilayer, the dissolution of material in a new solvent, or capturing the characteristic colors emitted from a library of fluorescent toxins.

1.2 Fundamentals of Computer Vision in Chemistry

1.2.1 Color Theory

Color theory is a crucial concept for chemists, enabling them to actively monitor reactions and observe changes over time. With the widespread availability of imaging devices, particularly in the age of social media, capturing image data from experiments has become more accessible. This convenience, however, brings the challenge of encoding color data accurately, a problem rooted in the rich history of color theory.

The foundation of modern color theory lies in the Young-Helmholtz theory, or trichromatic theory, developed in the nineteenth century.*[1] Initially proposed by Thomas Young, this theory suggested that the human eye has three types of photoreceptors. Hermann von Helmholtz later expanded on this idea, identifying these photoreceptors as sensitive to blue, green, and red light (more appropriately referred to as "short," "medium," and "long" wavelengths, given that the peak and distribution in response

*1 The socioeconomic significance of color is a richer story still, but not one that can be done justice in this book chapter.

for each receptor span more wavelengths than one specific color label). Tristimulus color theory is fundamental to understanding how colors are perceived and represented [2]. Several key concepts expand on this foundation, including color-matching functions, spectral reflectance, illuminants, and tristimulus values.

Color matching functions are critical in converting the spectral power distribution of light into tristimulus values. These functions, denoted as $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, describe the sensitivity of the human eye to different wavelengths of light. They are derived from experimental data on how humans perceive colors and are standardized by the International Commission on Illumination (CIE). Mathematically, the tristimulus values X, Y, and Z can be calculated using the color-matching functions described in Eqs. (1.1)–(1.3):

$$X = \int_{380}^{780} S(\lambda) \,.\, \bar{x}(\lambda), d\lambda \tag{1.1}$$

$$Y = \int_{380}^{780} S(\lambda) \,.\, \bar{y}(\lambda), d\lambda \tag{1.2}$$

$$Z = \int_{380}^{780} S(\lambda) \,.\, \bar{z}(\lambda), d\lambda \tag{1.3}$$

where $S(\lambda)$ is the spectral power distribution of the light source. This, in turn, relates to spectral reflectance, which is a measure of how much light is reflected by a surface at each wavelength. It is defined as the ratio of reflected light to incident light and varies with the wavelength. The spectral reflectance function, $R(\lambda)$, characterizes the color of a surface, in a relationship with the spectral power distribution of the light source (or illuminant), $I(\lambda)$, according to Eq. (1.4). The visual intuition for arriving at XYZ color values is shown in Figure 1.2.

$$S(\lambda) = I(\lambda) \,.\, R(\lambda) \tag{1.4}$$

The color spaces, several of which are introduced below and first hinted at in Table 1.1, allow us to conceptualize colors in an exploitable numerical framework (think back to the dress conundrum). Understanding and utilizing these color spaces is essential for accurately capturing and interpreting the color data in chemical case

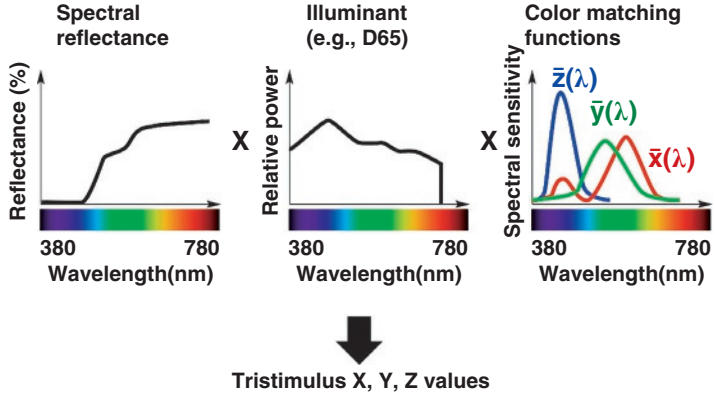

Figure 1.2 Visual representation of the calculation of the tristimulus XYZ values.

studies in later sections of this chapter and beyond. Up front, there are two common factors of color spaces that cater toward intuitively quantifying what a human observer may perceive perceptual uniformity and visualization.

Perceptual uniformity means that equal distances between colors in the color space correspond to equal perceived differences in color. In other words, if two colors are separated by the same amount in this space, they will look equally different to the human eye. This makes the color space more consistent with how we actually perceive colors, even though it might not be a simple linear mathematical relationship. A linear space means that the points in the color space (i.e., the vectors) can be added and multiplied (or scaled) correctly and will result in another color in that space. When taking the Euclidean distance between two points in a color space, perceptual uniformity means that the distance is the same as the difference in how a human perceives color. Visualization (in this context) means the ability of a human observer to look at the individual elements of the color vector in a color space and understand what the element represents. These two factors account for a color space that is easily interpretable by a human but are not necessary for a color space to be functional (e.g., RGB spaces, while not having these factors are commonly used in capture devices).

1.2.1.1 RGB Color Spaces

The RGB color spaces (linear RGB, sRGB, Adobe RGB, and more) are among the most widely used color models, especially in digital imaging, computer graphics, and electronic displays. These models are based on the additive color theory, where colors are created by combining different intensities of red, green, and blue light. The additivity of red, green, and blue light, along with exemplar 8-bit color coding in RGB, is shown in Figure 1.3.

Certain RGB color spaces are device-dependent, meaning that the colors displayed can vary across different devices due to variations in hardware characteristics, such as screens or printers. Examples of such device-dependent spaces include those used in monitors, cameras, or printers. However, there are also device-independent RGB spaces, like sRGB, which aim to ensure consistent color reproduction across different devices by standardizing the color transformations. While this variation among devices may seem to initially be a major drawback of using RGB, the ability to transform between color spaces leads to color models that mitigate such risk. RGB space itself is non-intuitive due to the difficulty of understanding how the R, G, and B components combine (or add) to produce the final color. It can be linear; however, there are also non-linear variations, such as the sRGB color space [3].

1.2.1.2 CIE XYZ and xyY Color Space

In the XYZ space, the X and Z values refer to chromaticity and Y luminance. Chromaticity defines what color something appears to be (its hue and saturation), while luminance defines how bright or intense that color appears. As demonstrated in Figure 1.4, objects with different Y (luminance) values might look like the same color but with different brightness levels, while objects with the same Y value but different X and Z values might have the same brightness but appear as different colors (chromaticity).

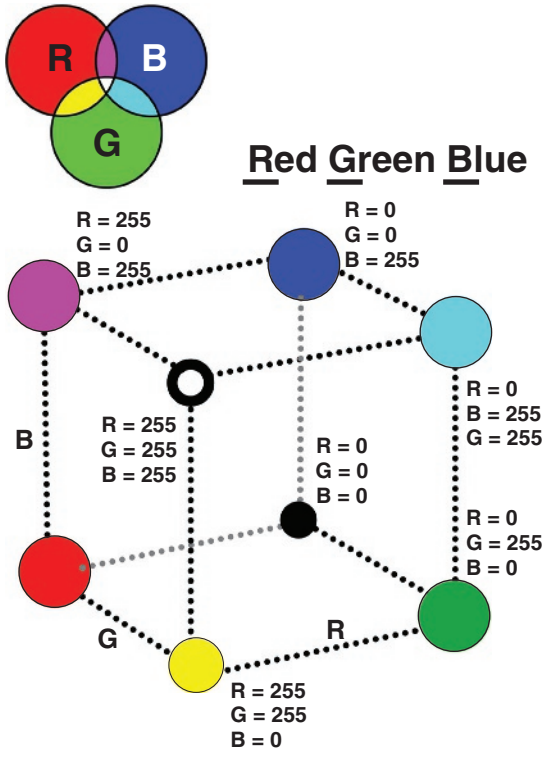

Figure 1.3 Top left: a demonstration of the additive properties of overlapping red (R), green (G), and blue (B) lights. Bottom: an overview of exemplar RGB values and their geometric positioning.

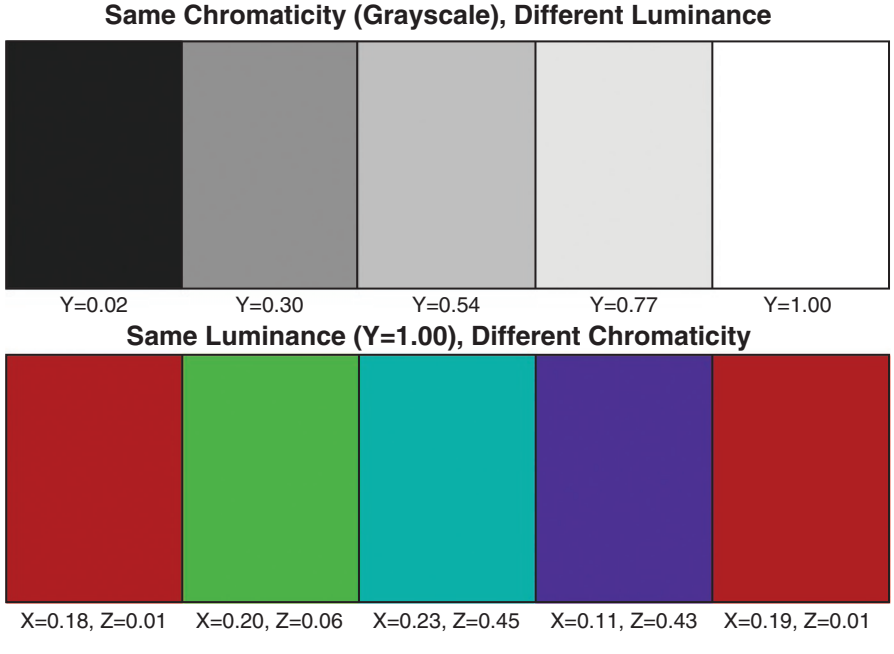

Figure 1.4 A simple demonstration of the distinction between chromaticity (X and Z values) and luminance (Y value).

Unlike RGB, this color space was designed to be device-independent and, as such, is often used as an intermediate color space as every color space has an appropriate transformation. For example, the transformation between XYZ and sRGB is given by Eq. (1.5) [4].

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{1.5}$$

In the equation above, the 3×3 matrix contains all the coefficients for transforming to RGB. These coefficients are empirically determined by correlating white in both spaces using various samples.

In a transformation closer to XYZ itself, we find the xyY color space, which is designed to separate chromaticity (x,y) from luminesce or brightness (Y). The conversion is accomplished according to Eqs. (1.6) and (1.7) (luminance in Y remains unchanged from the XYZ space).

$$x = \frac{X}{X + Y + Z} \tag{1.6}$$

$$y = \frac{Y}{X + Y + Z} \tag{1.7}$$

As we will see in the case studies later, this transformation is a useful example of being able to decouple color from lighting information.

1.2.1.3 Hue, Saturation, and Value

This color space defines hue (H), saturation (S), and value (V). Hue is itself what determines the color, while saturation describes how non-white the color is. Higher saturation means more vivid colors. Value is the brightness of a color, with its extremes being black and white. The HSV color space is a non-linear transformation of the RGB color space, expressed in polar coordinates. It is also device-dependent. This non-linearity makes it harder for human perception due to not being perceptually uniform.

The relationship between HSV and RGB can be expressed according to Eqs. (1.8)–(1.10).

$$V = \max(R, G, B) \tag{1.8}$$

$$S = \begin{cases} 0 & \text{if } V = 0 \\ \dfrac{V - \min(R, G, B)}{V} & \text{if } V > 0 \end{cases} \tag{1.9}$$

$$H = 60^\circ \times \begin{cases} \dfrac{G - B}{V - \min(R, G, B)} & \text{if } V = R \\ \left(2 + \dfrac{B - R}{V - \min(R, G, B)}\right) & \text{if } V = G \\ \left(4 + \dfrac{R - G}{V - \min(R, G, B)}\right) & \text{if } V = B \end{cases} \tag{1.10}$$

In Figure 1.5, the HSV visualization has assumed that S and V are normalized on a percent scale of 0–100, relative to the 8-bit RGB scale.

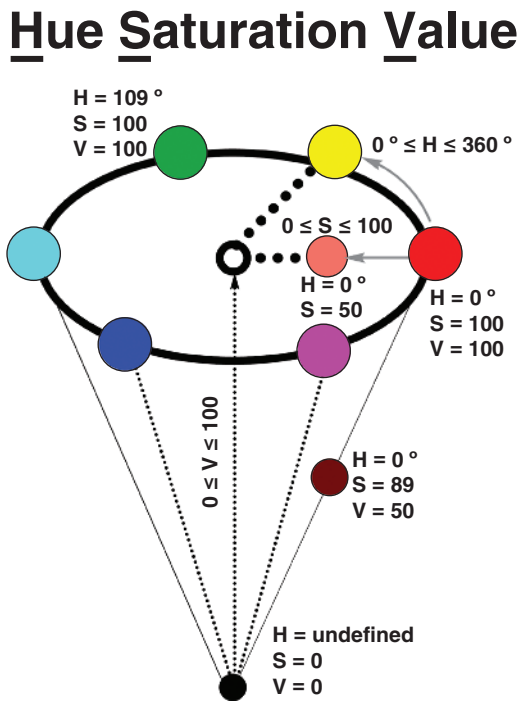

Figure 1.5 A cone-shaped representation of the HSV color space, showing representative HSV values for illustration. Cylindrical representations of this space also exist. While the cone is more intuitive, representing decreasing color saturation as value decreases, the cylindrical variant of this model is also useful, especially for computational reasons.

$$Normalized\ R\ or\ G\ or\ B\ = \frac{R\ or\ G\ or\ B}{255} \times 100 \tag{1.11}$$

1.2.1.4 CIE–L*a*b*

The CIE–L*a*b* color space is a powerful tool for color representation, providing a valuable attempt at perceptually uniform space. The elements of this color space represent lightness (L*), red to green (a*), and yellow to blue (b*). The L*a*b* color space was initially designed in 1976 to be a proxy for human vision (Figure 1.6). The distance between two colors correlates to the difference in the human perception between those two colors. This property makes L*a*b* attractive for marrying by-eye perception to a framework of color that leads to an intuitive interpretation of the numbers. This is not the case for RGB, for example. Linking the CIE–L*a*b* color space with human biology, the chromatic axes, a* and b*, link to two modes of human color blindness, namely *deuteranopia* (red-green blindness) and *tritanopia* (blue-yellow blindness) [5, 6].

While the more involved transformations to get to the CIE–L*a*b* space from XYZ and RGB are left for consideration elsewhere [7], a calculable consequence of this particular model is worth highlighting, namely ΔE. Described in Figure 1.6, Eq. (1.12), and Table 1.2, the simplest definition of ΔE is as the Euclidean (straight line) distance through the color space. This metric enables one to calculate the magnitude of contrast between two colors in a way that is agnostic to the colors involved.

$$\Delta E_{1976} = \sqrt{(L_2 - L_1)^2 + (a_2 - a_1)^2 + (b_2 - b_1)^2} \tag{1.12}$$

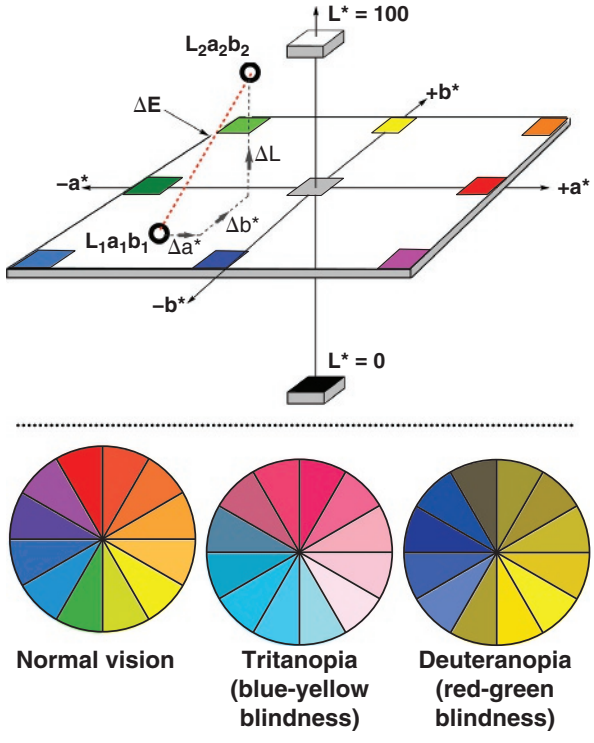

Figure 1.6 Top: A visual representation of the CIE–L*a*b* color space. Bottom: Color wheels comparing unhindered human vision versus two modes of color blindness.

Table 1.2 Interpretation of the magnitude of ΔE color contrast metric.

Δ*E*	Interpretation
0	Change is not perceptible by the human eye.
1–4	Perceptible change with very close observation.
5–10	Subtle change, perceptible at a glance.
11–49	Color changes are more similar than the opposite.
50–99	Colors changes are increasingly obvious.
100+	Color changes are stark.

From these theoretical frameworks of color, we are now ready to explore how these data are likely to be captured in the first place.

1.2.2 Digital Photography Basics

1.2.2.1 How Does a Digital Camera Work?

Before digging the hole down to image analysis, algorithms, and computer vision in chemistry, we must first understand the picks and shovels we are using to do the digging. For us, we need to understand how a digital camera works. The nuts

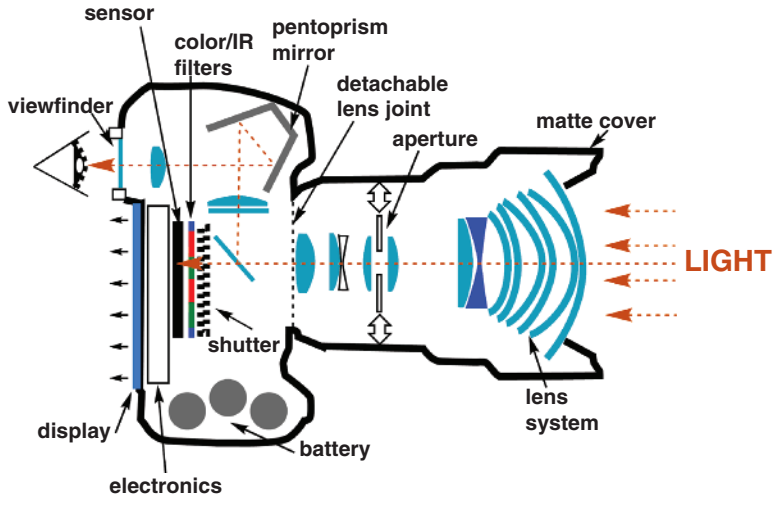

Figure 1.7 A simplified top-down view of the inner workings of a DSLR digital camera.

and bolts of a traditional digital single-lens reflex (DSLR) camera are shown in Figure 1.7.

When you take a photograph of the by-eye visible chemical phenomenon you want to capture – catalytic or not, debated dress or otherwise – you have begun the process of turning subjective interpretation to objective measurement amenable to ML. First, light from the chosen light source (the fluorescent lights of your fume cupboard, a light box filled with LED strips, and the sun through the laboratory window) will be reflected, scattered, and transmitted from the object of interest. The resulting light rays will be focused using the camera's lens system, wherein the light path is made to converge on the image sensor. Before the light ever reaches the photosensitive sensor itself, it first passes through two filters:

1. An infrared filter, cutting out wavelengths greater than those in the red end of the visible spectrum.
2. A color (typically Bayer) filter that enables color information to be derived from light hitting the sensor. In most cases, the color filter will have an arrangement of red, green, and blue components for which the green is represented twice as often as the other filters. This relates to our human light sensitivities, which are greater for wavelengths in the green region than elsewhere.

Having passed through these filters, the remaining light will trigger reactions on the semiconductor material that makes up the sensor itself. Electronically, the sensors are either CCD or CMOS in design, the deeper distinctions between which are beyond the concerns of this chapter.

The analogue signal collected is then sent to an analogue-to-digital converter, which then enables the data to be processed for the likes of image compression and noise reduction. The resulting image is then temporarily stored in a memory buffer

before being sent to the long-term storage on the device. After all of this, the image can then be retrieved for display on a camera or computer display.

Ultimately, given how the data are collected in a digital camera, the chemist seeking to exploit computer vision for their research purposes must always be aware that the images created will be affected by:

1. The light source used to illuminate the chemical reaction.
2. The object composition of the system under study (e.g., matte or gloss).
3. The angle between the incident light and the camera that serves as our detector.
4. The lens choice (in cameras where this can be interchanged to alter the field of view, depth of field, and related photographic considerations).
5. The camera type and manufacturer (e.g., full frame vs. smaller sensor sizes).

Practically speaking, rather than the paradox of choice trapping us into sinking time picking the perfect camera, it is often more important to apply a consistent set of photographic conditions within any one application or piece of research. Such an approach allows maximum interpretability of chemistry outcomes with minimal risks of comparing images collected under different environmental conditions. It is the digital format of images and videos captured in the memory storage of the digital cameras that ties their application in computer vision to our earlier discussions of color theory.

1.2.2.2 Image and Video: The "Crude Oil" that Fuels Computer Vision

Be it a JPEG or PNG file for image analysis or MP4, AVI, or MOV for video analysis, the base currency of visible range computer vision applications in chemistry is the three-dimensional array of RGB values. These RGB arrays capture the triplet of red, green, and blue values associated with each pixel of the image. In relation to each pixel, every position in a RGB array is composed of three values, running from 0 to 255, or 256 integer possibilities total. These numbers, far from arbitrary, are based on the binary (base 2) numbers. Oftentimes, colors are stored as 8-bit binary numbers ($2^8 = 256$). While each "slice" of the RGB array can be accessed and used in isolation (Figure 1.8), it is also commonplace to reduce the array dimensionality by converting the RGB values to single grayscale values [8]. A single grayscale value is commonly calculated as a linear combination from RGB and, like the Bayer filter featured in a camera's anatomy (Figure 1.6), green values are higher weighted than R or B, tipping the mathematical hat to our biological predisposition towards higher sensitivity to green wavelengths than to any other.

$$Gray = 0.21R + 0.72G + 0.07B \tag{1.13}$$

Video, as opposed to image analysis, adds the dimension of time, where a timestamped list of such RGB arrays comes into play. The higher the camera's image resolution, the more pixels there are capturing RGB values, and the larger the resulting data array to be read in for programmatic analysis and ML. Higher image resolution tends to be more important in microscopic applications, where finer details may need to be segmented. Where video is concerned, the higher the available frame rate (or frames per second, FPS) of video capture, the more frames of RGB arrays can be collected within a given recording time. Higher temporal resolution opens opportunities to track faster processes and harness the power of slow-motion videography.

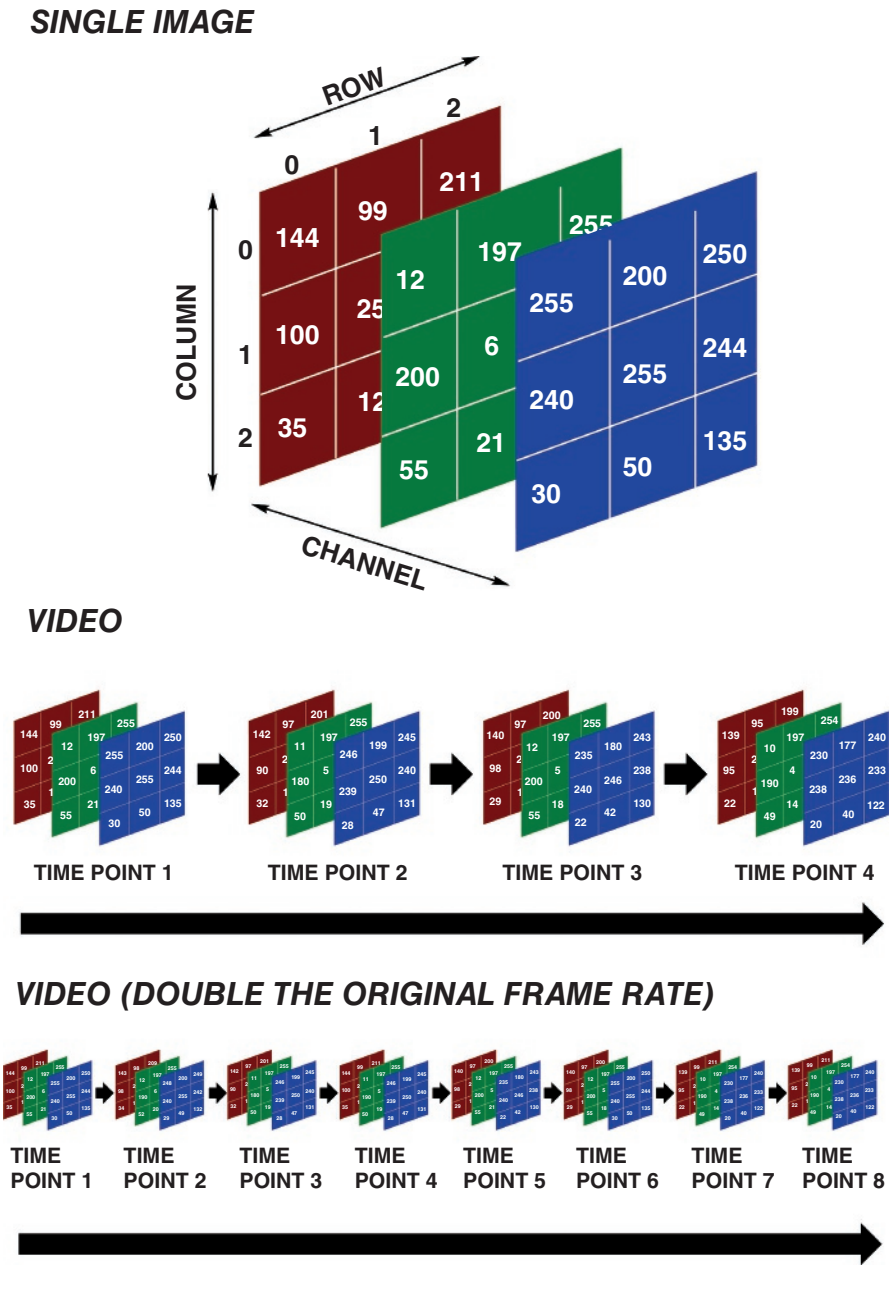

Figure 1.8 Top: a visual representation of the RGB array data structure. Bottom: a simplified overview of how RGB arrays can represent video frames for time-resolved analysis.

The management of larger and larger data inputs for computer vision analysis demands that issues of software performance, memory management, and (big) data distribution be addressed, but those matters will be left for a more computer

science-focused discussion. Our focus here remains on chemistry and catalysis applications of computer vision.

1.2.2.3 Algorithms and Computer Vision

With an appreciation for common color models, photography, and important data structures for programmatic analysis, one is ready to understand some common algorithms applied in computer vision. From here, applications in chemistry are then easier to identify and categorize.

1.2.2.4 Color Distributions/Histograms

In the context of computer vision, color histograms are a graphical representation of the distribution of colors in an image. They quantify the number of pixels possessing each color intensity level, typically in each of the RGB (red, green, blue) channels. This is achieved by scanning through each pixel in the image, determining its color value, and then incrementing the corresponding bin in the histogram. The final histogram represents the frequency distribution of the colors in the image. The grayscale (or luminosity) histogram provides a visual representation of the linear combination of the R, G, and B quantities, according to Eq. (1.13). These histograms are useful for a variety of reasons. They provide a compact summary of the color content of an image, making them valuable for image retrieval, object recognition, and image classification tasks. We will see examples of their application in Section 3. See an exemplar comparison of R, G, B, and luminosity histograms in Figure 1.9.

1.2.2.5 Texture Analysis

In the context of computer vision, texture analysis involves examining the spatial distribution of pixel intensities within an image to identify and quantify patterns, textures, and structures [9]. This process helps distinguish regions in an image based on their surface characteristics rather than their color or intensity alone. Texture analysis is calculated using various statistical, structural, and spectral methods. One of the simplest and perhaps most intuitive of such metrics is derived from pixel thresholding. The contact metric (Figure 1.10) measures the total perimeter outlined by boundaries between the pixels above and below the set threshold. A low value indicates few or no boundary regions, few pixels distinguished on the basis of the threshold, and, therefore, a short overall perimeter length. Such low contact values would be indicative of a more homogeneous texture or smooth distribution of color over an image. Conversely, there are more instances of black pixels meeting white, a larger calculated perimeter, and a resulting suggestion of a more heterogeneous image composition [10].

Going deeper, more intricate methods of calculating texture involve analyzing these pixel value distributions and relationships within defined windows or the entire image. For example, the gray level co-occurrence matrix (GLCM) quantifies the number of instances of pixel value pairs in a specified spatial relationship in an image (Figure 1.11) [10–13]. It is used to extract various statistical measures that describe the texture of the image in question. Such measures include contrast, angular second moment (ASM), and entropy, and are defined in Eqs. (1.14)–(1.16) below.

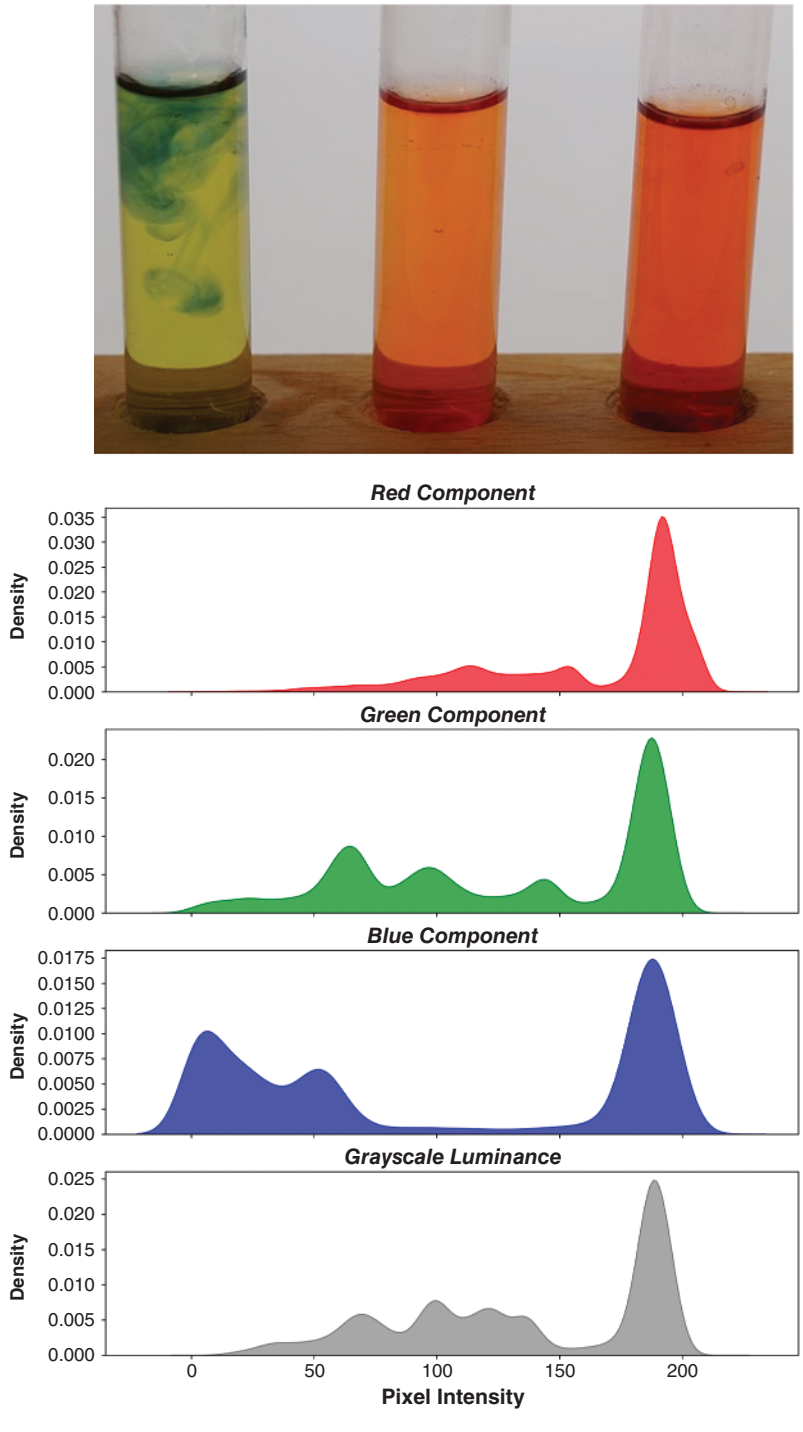

Figure 1.9 Top: three test tubes filled with three differently colored liquids. Bottom: a demonstration of the R, G, B, and luminance histograms summarizing the ensemble composition of pixel colors captured in the top image. *Source*: B. Lachner/ Wikimedia Commons/Public domain.

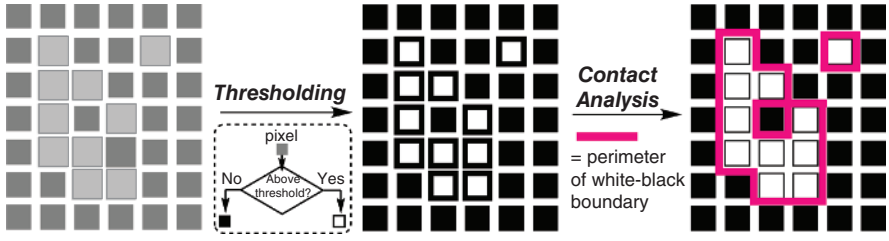

Figure 1.10 Conceptual representation of the contact analysis used for mixing analysis.

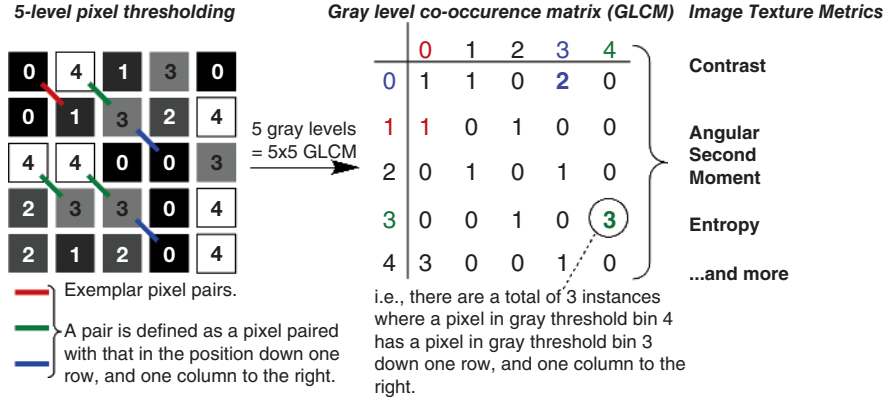

Figure 1.11 Conceptual overview of the construction of the gradient co-occurrence matrix (GLCM).

$$Contrast = \sum_{i,j}(i-j)^2 . P(i,j) \tag{1.14}$$

$$Angular\ Second\ Moment\ (ASM) = \sum_{i,j} P(i,j)^2 \tag{1.15}$$

$$Entropy = -\sum_{i,j} P(i,j) \,.\, \log\left[P(i,j)\right] \tag{1.16}$$

where $P(i,j)$ is the value of the element in row i, column j of the GLCM, normalized as a fraction of the sum of all values in the matrix (the grand sum).

$$P(i,j) = \frac{a_{ij}}{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} a_{ij}} = \frac{a_{ij}}{grandsum(GLCM)} \tag{1.17}$$

$$\sum P(i,j) = 1 \tag{1.18}$$

To understand the relative (high and low) values of texture metrics, consider the example texture analysis shown in Figure 1.12. Our by-eye comparison of the test tubes leads to the intuition that Tube 1 is more poorly mixed or more heterogeneous than Tube 2. In other words, the color distribution is more uniform in the second test tube. From the exemplar texture metrics displayed in the accompanying graph, we can quantitatively state that the color distribution in Tube 1 contains

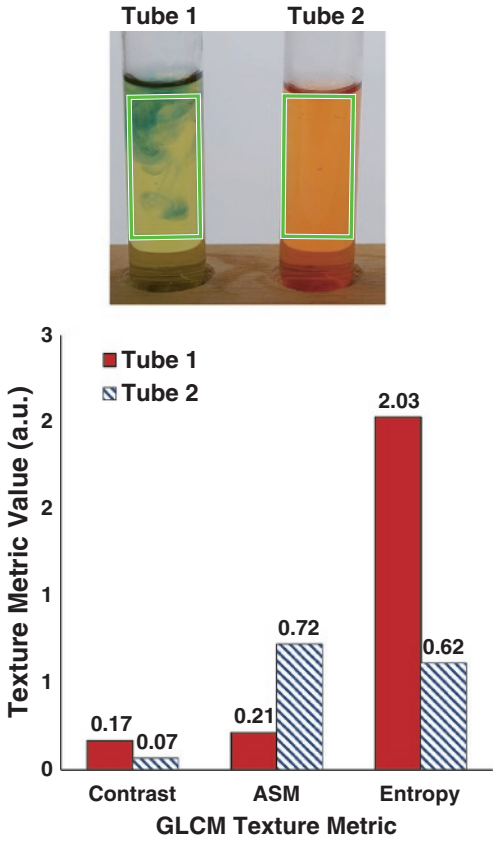

Figure 1.12 Exemplar calculations of GLCM texture metrics for the green highlighted ROI in each of two test tubes.

more local variation (higher contrast), contains more uniformity or patterned regularity (lower angular second moment, or ASM), and is closer to maximum randomness (higher entropy) versus Tube 2. All these metrics are derived from the GLCM of the region of interest selected within each test tube from the image file.

These exemplified analyses represent one class of image segmentation, classifying pixels above and below certain thresholds, and not always binary. We can take the idea of segmentation further to understand how images can be transformed to extract (or eliminate) certain features on the road to more complex machine-learning efforts.

1.2.2.6 Convolution and Edge Detection

Segmentation and object detection are crucial techniques in computer vision. With countless colors, textures, phases, vessels, and processes, the chemistry applications of computer vision are no exception [14]. As we have seen most clearly with the Contact metric defined above, segmentation involves partitioning an image into distinct regions to isolate objects or boundaries, thereby simplifying image representation. While not limited to the following, additional methods to achieve segmentation include edge detection and advanced deep learning approaches using convolutional neural networks (CNNs). Common to both of these methods is a need to introduce the last fundamental computer vision concept we will consider before

discussing chemistry- and catalysis-specific applications of computer vision and downstream ML, *convolution*.

Image convolution is a fundamental operation in computer vision used to transform images and thus extract features and later image quality. It involves applying a kernel (or filter) to an image, where the kernel is a small matrix of numbers that is slid over the image, performing element-wise multiplication and summing the results to produce a new pixel value (Figure 1.13). This process is repeated for every pixel in the image, resulting in a transformed output image. Different kernels have distinct structures and impacts on the image, as demonstrated in Figure 1.14.

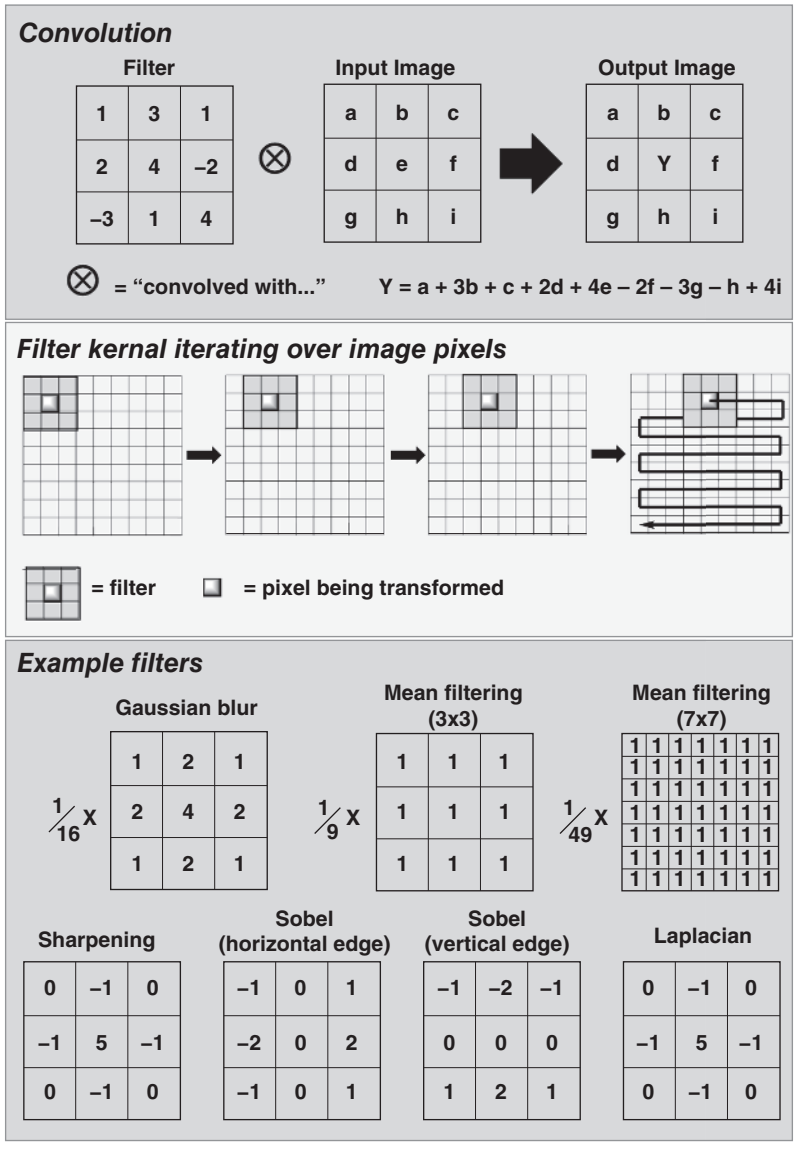

Figure 1.13 A simplified overview of convolutional operations used for image transformations.

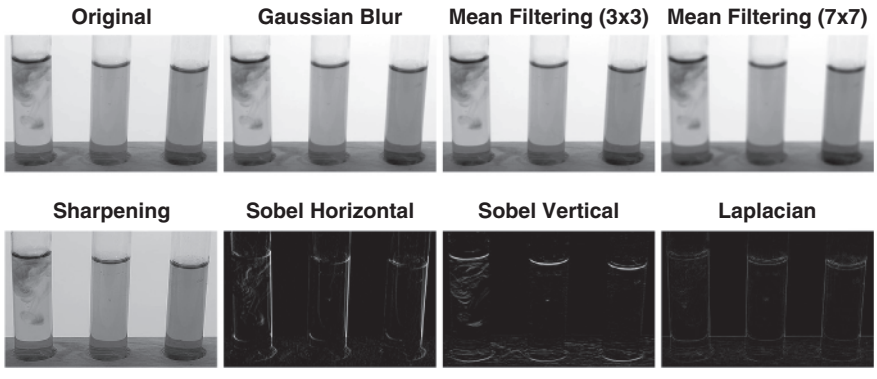

Figure 1.14 Exemplar demonstrations of the effect of using the convolutional filters described in Figure 1.13 on the output image.

From color spaces and their interconversions to photography hardware, color arrays, and all the valuable manipulations thereof, this combined software and hardware approach to understanding computer vision leaves us well-placed to consider the illustrative chemistry case studies below. Considering the broader topic of ML, we learn that, with computer vision, it is as much about the choice of machine as it is about any of the underlying software that comes front of mind with a term like ML.

1.3 Computer Vision and Machine Learning in Chemistry

Given that, at the time of writing, the applications of computer vision in catalysis itself remain sparse, the examples below include catalysis and phenomena that, while not directly related to catalysis, represent analytical challenges that will likely be relevant to future applications in the field.

1.3.1 Single Image Applications

1.3.1.1 High-throughput Digital Fingerprinting of Drug-like Compound Libraries

In 2011, Hodder and coworkers combined the hardware and software needed to build a system for assessing the quality of a library of drug-like compounds solvated in dimethyl sulfoxide (DMSO) [15]. Their system exemplified several key features that make computer vision valuable to chemistry and, potentially, catalysis. The team's use of a telecentric lens system was key to enabling distortion-free imaging of an entire well plate without optical distortion of parallax errors leading to unequal imaging of central versus outer wells. A telecentric lens is a specialized type of optical lens designed to have constant magnification regardless of the distance between the lens and the object being imaged. This unique characteristic is achieved by ensuring that the chief rays (rays passing through the center of the aperture) are parallel to the optical axis in either object space, image space, or both.

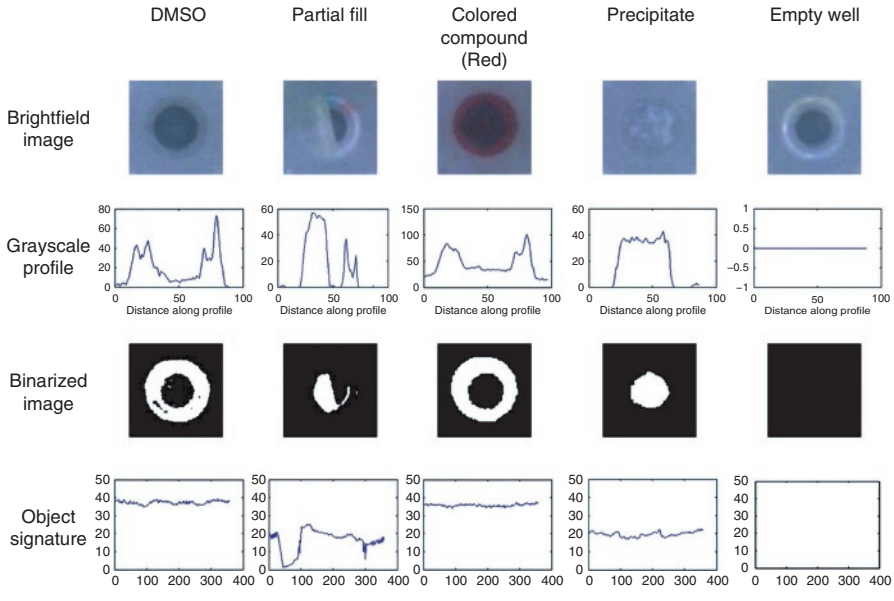

Figure 1.15 Use of grayscale profiling and binary image thresholding to characterize solubility behavior of a drug-like compound library. Reproduced with permission from [15]/SAGE.

From a chemical perspective, the solvation screening of the available compound library involved colored compounds, colorless compounds, soluble and precipitous insoluble candidates, as well as empty wells and solvent-only wells for calibration. With such a range of physicochemical behaviors, a range of visible "artifacts" could be captured and processed with computer vision. Using the grayscale lightness histogram distribution from a single well image, it was possible to create a binary black and white image to visually exaggerate the homo- or heterogeneity of the DMSO-solvated sample. The team used a combination of visible and infrared imaging in order to more fully distinguish visibly similar wells, especially in cases where colored compounds masked underlying precipitates. From here, a unique compound "signature" (the details of which are not clear from or not disclosed in the publication), providing a digital fingerprint with which that compound could be stored in a database containing all computer vision outputs (Figure 1.15).

1.3.1.2 Segmentation and Parallel Analysis of Nerve Agents

In 2018, Anslyn and Marcotte reported the use of photographic methods to measure fluorescence from self-propagating cascades used to amplify the signal and thus optically detect ions of various nerve agents (Figure 1.16) [16]. While the measurement of fluorescence for such purposes typically requires sophisticated, capital-intensive methods like fluorescence microscopy, UV-vis spectroscopy, and X-ray photoelectron spectroscopy, a demand for a more accessible and field-deployable technique was required in this case. To this end, a smartphone was used as the optical detector of fluorescence for samples housed in an opaque stage built from black toy building blocks.

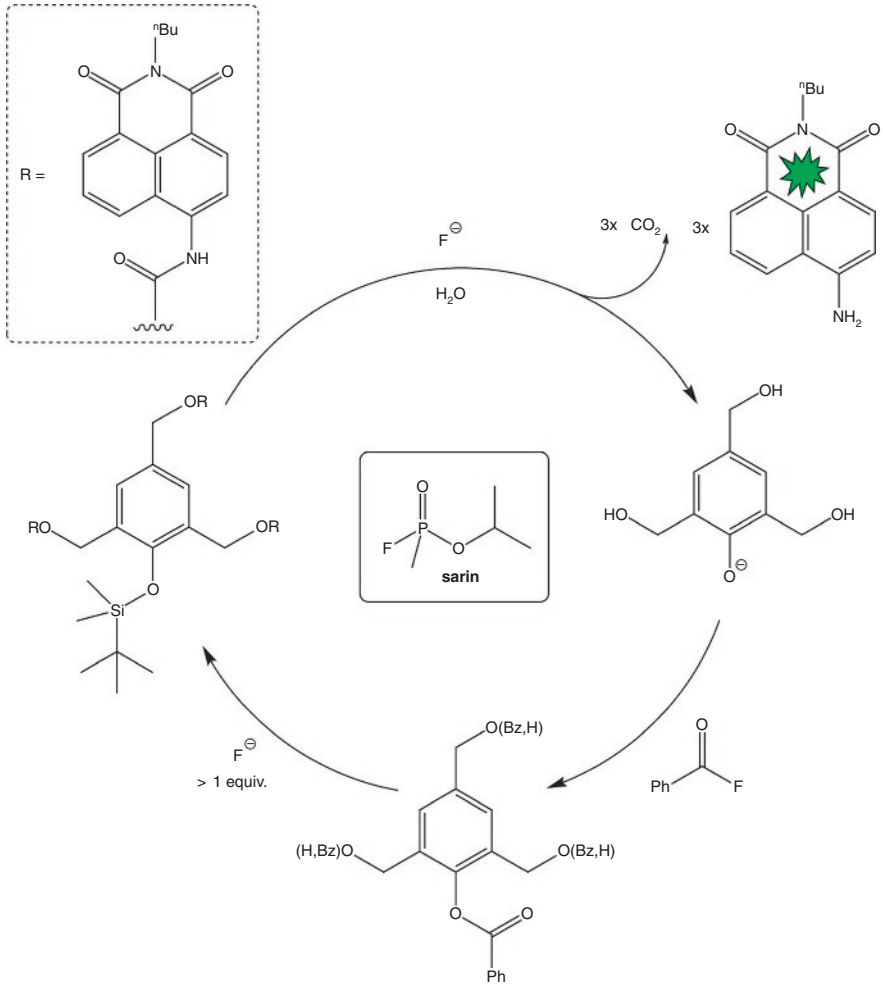

Figure 1.16 A simplified overview of one of two fluorescence amplification cascades used to generate samples amenable to high-throughput imaging of fluorescent samples via a computer vision approach. Adapted from [16].

Two elements of this work are important from a computer vision perspective, and both appear in Figure 1.17. First, the team employed the xyY color space, closely related to the above-mentioned XYZ color space (see Section 2.1.2). In doing so, the team used the *chromaticity* components, x and y, separate from the lighting-focused luminosity parameter, Y, to find the representative color of each sample well. In relation to the above-mentioned computer vision analytics and algorithms, the representative chromaticity of each sample was calculated based on the median pixel distribution within a given sample's color histogram for components x and y. By discounting the luminosity component, it was envisaged that the method would remain robust to changes in mobile phones used, owing to the inevitable variation in camera sensors when moving from one device to another. Plotting calibration samples on the xy-plane, the team interpolated the concentration of unknown nerve

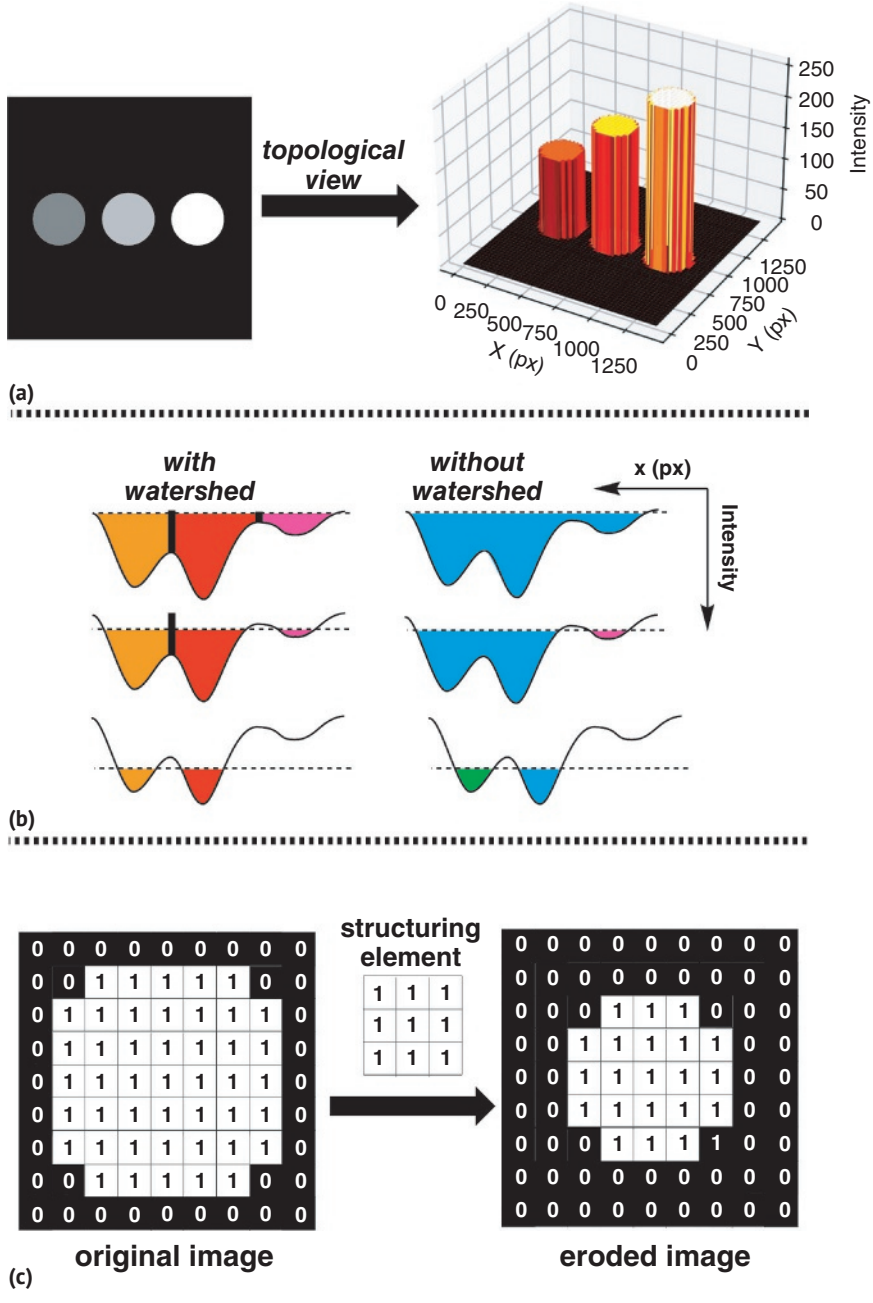

Figure 1.17 (a) Top: a demonstration of how grayscale values on an image can be viewed as a topological relief. (b) Middle: a one-dimensional pixel representation of filling the topological relief, where the bold black "dams" represent the identified watershed segments between local minima. (c) Bottom: an illustration of the morphological erosion technique employed in the second part of the computer vision algorithm by Anslyn's team.

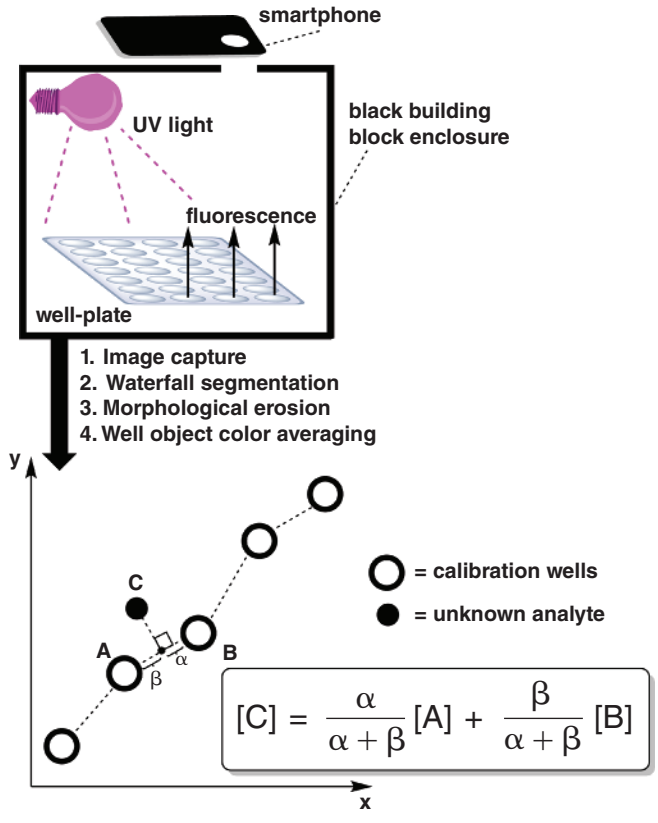

Figure 1.18 Top: A representation of the photographic hardware and staging used for the fluorescent imaging of nerve agents. Bottom: The resulting statistical framework, built around the xy-plane of the xyY color space, through which unknown sample concentrations could be estimated from the extracted xy-color parameters of calibration wells.

agent samples by using a piecewise linear regression approach. Here, Euclidean distances (i.e., lines between two points) were calculated for every adjacent pair of data points. Unknown sample concentrations were then estimated according to the equation embedded in Figure 1.18.

Yet, beyond the application of color theory and simple statistical analysis to produce a field-deployable method, it is how the individual samples on the 96-well plate were identified that merits further illustrative discussion. To this end, the team applied the so-called *waterfall* algorithm (Figure 1.17a,b) to segment one well from another and thus automatically optimize the selection of pixels used for the above-mentioned calculation of median chromaticity in each case. The waterfall algorithm is a refinement of the watershed algorithm, treating the grayscale image as a topographic surface, where the pixel values represent elevation. It simulates the process of water flooding from the lowest points (local minima), progressively filling up basins, and defining boundaries at points where water from different

basins would meet (Figure 1.17b). This method effectively segments the image into distinct regions based on the topography of the intensity landscape. However, as noted by Anslyn and Marcotte, the watershed algorithm can be sensitive to noise and over-segmentation, creating too many small regions. The waterfall algorithm addresses this issue by iteratively merging regions based on certain criteria, such as similarity in intensity or connectivity, to produce more meaningful and less fragmented segments. Essentially, the waterfall algorithm enhances the watershed method by introducing a post-processing step that reduces over-segmentation, leading to more coherent and robust image segmentation outcomes. The overall practical impact of this transformation is to shrink the object whose remaining pixels are then used for further analysis (Figure 1.17c). One can imagine such methods being applied to the analysis of high-throughput catalyst screening samples.

1.3.1.3 Neural Networks for Identifying Glassware

While case-specific edge detection, object recognition, and segmentation approaches are established aplenty in computer vision, the identification of myriad variations on chemistry glassware and its multiphase contents represents a notable challenge in chemistry-focused computer vision applications. In 2020, the Eppel and Aspuru-Guzik groups collaboratively released an approach to this problem, and in so doing provided an ideal exemplar of using CNNs in chemistry-centered computer vision [17].

CNNs are a class of deep learning models specifically designed for and trained on tasks involving image analysis and computer vision. They are inspired by the organization of the animal visual cortex, where individual neurons respond to stimuli only in their receptive fields, allowing for hierarchical feature extraction. In CNNs, convolutional layers perform feature extraction by applying a set of learnable filters (kernels) to input training images, first discussed in Section 2.2.3 (Figures 1.13 and 1.14). These filters detect various features such as edges, textures, and patterns, preserving spatial relationships through weight sharing and parameter sharing. Pooling layers then down sample the feature maps, reducing computational complexity and helping the network become more robust to translation and distortion invariance. Through multiple convolutional and pooling layers followed by fully connected layers, CNNs can learn hierarchical representations of visual data, enabling them to perform the likes of the object and chemical phase detection tasks reported by the Eppel and Aspuru-Guzik groups. The fundamental connectivity of a deep neural network and the distinct elements of CNNs are shown in Figure 1.19.

Using a labeled library of 2187 images of reaction vessels containing different colors and phases of material, the team employed several different CNNs to investigate how best to segment the reaction vessel from the material within. It is worth noting that the image library used for training in this case was relatively small, especially compared to available training libraries like ImageNet, which, at the time of writing, contains over 14 million openly available training images [18]. Among the methods tested was the R-Mask CNN approach [19]. An R-mask CNN (region-based mask convolutional neural network) is a type of CNN architecture specifically designed for semantic segmentation tasks in computer vision. It combines the

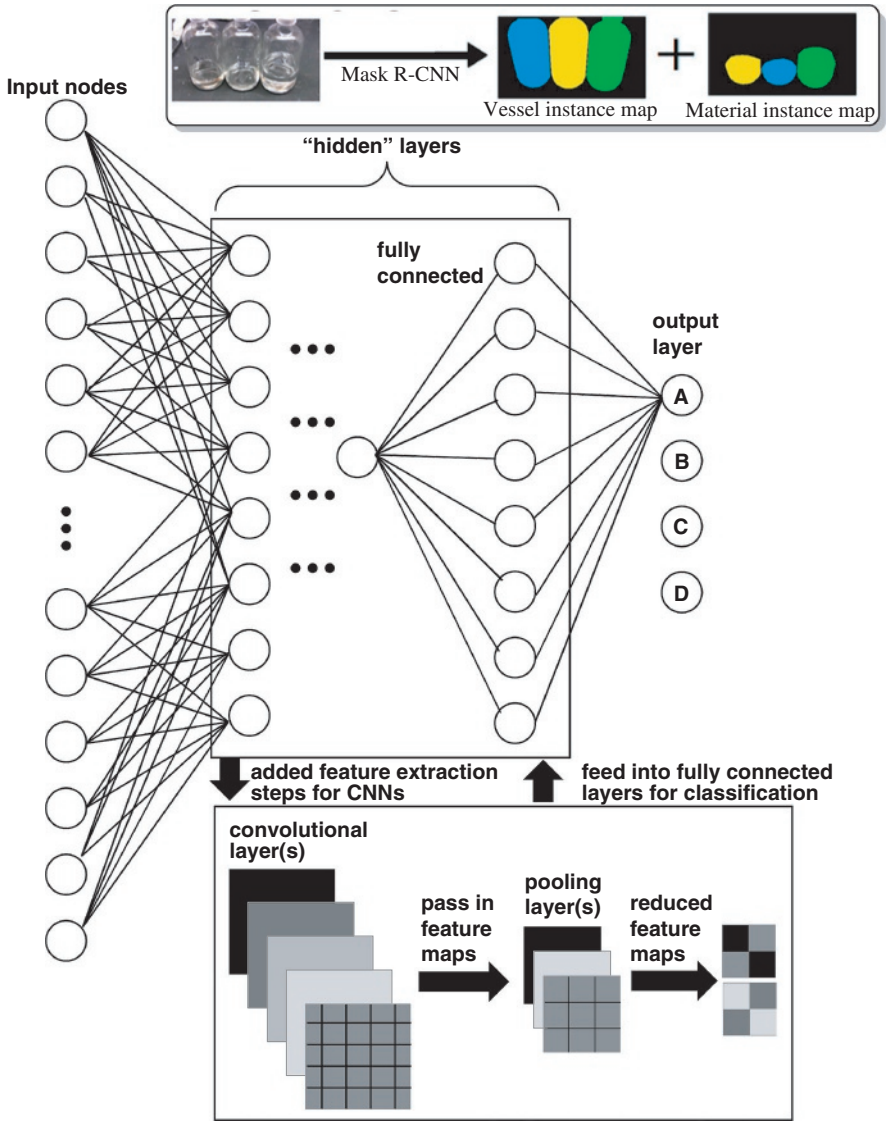

Figure 1.19 An overview of deep neural networks and the layers through which CNNs represent a strategic architecture for the analysis of grid-like or image-based input data. Inset: An example of the output segmentation of liquid from the vessel, published by Aspuru-Guzik. Reproduced with permission from [17]/American Chemical Society.

strengths of region-based convolutional neural networks (R-CNNs), which are adept at object localization, and mask-based CNNs, which excel at pixel-level segmentation. In R-Mask CNNs, the network first generates region proposals using techniques like selective search or region proposal networks (RPNs). Then, for each region proposal, the network simultaneously predicts the class label and generates a binary mask indicating the object's presence at the pixel level. This allows for accurate object detection and precise segmentation in images, making R-Mask CNNs

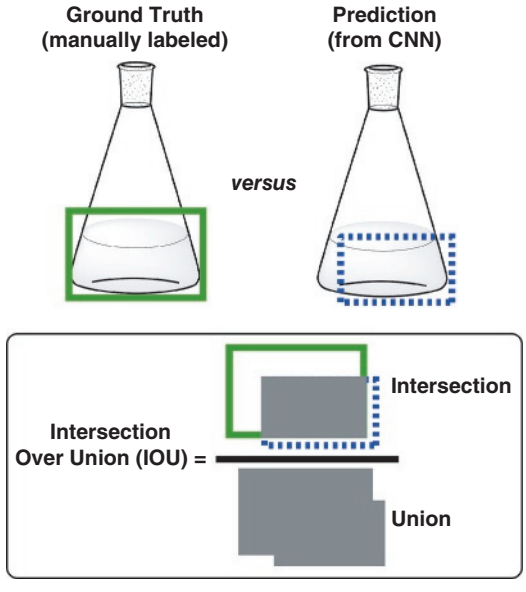

Figure 1.20 A visual representation of the IOU evaluation metric is applied to understanding CNN performance versus human-labeled definitions of object detection.

particularly suitable for tasks such as instance segmentation, where distinguishing between individual object instances is crucial.

Based on the chosen scoring systems of the CNNs applied, it was found that segmentation between the vessel and a solid or liquid material inside was higher scoring than more granular segmentation of multiphase systems (e.g., liquid vs. foam). Indeed, this point helps exemplify a broader point of importance when considering CNNs and ML for image analysis. CNNs can make various types of errors in predictions, such as false positives (where the model incorrectly identifies a pattern or feature as present) and false negatives (where the model fails to identify a pattern or feature that is present). These errors are manifested in predicted values versus actual values, often expressed as metrics like mean squared error (MSE) or root mean squared error (RMSE). In the Eppel and Aspuru-Guzik study [17], the more CNN-focused evaluation metric known as intersection over union (IOU) was applied (Figure 1.20). The roots of errors in CNNs can include the likes of overfitting, which occurs when the CNN model learns noise or irrelevant details from the training data, leading to poor generalization of new data. Conversely, if the CNN model is too simple, it may not capture the complexity of the chemical data, leading to high bias and systematic errors. Additionally, errors can arise if the training data (in this case, labeled images) contains noise, mislabeled samples, deliberate adversarial manipulation [20], or is not representative of the broader chemical space. Having said this, if errors themselves can be quantified, they can be used as an additional model parameter to help improve CNN performance [21].

Such single image-based applications of computer vision are far more broadly precedented than is detailed here [7, 22–24]. However, going beyond images into the domain of video analysis represents a relatively more opportune and underdeveloped area for catalysis and ML. This brings us to the final application-focused section of this chapter.

1.3.2 Video Analysis Applications

When we move from analyzing static images to dynamic video data, we build in the dimension of time, opening ML opportunities in time series analysis for reaction monitoring, mechanistic analysis, and a more characteristic level of problem diagnosis than is possible from single image inputs.

1.3.2.1 Phase Separation and Partition Coefficients

An emerging area of time-resolved computer vision for chemical interests in analyzing the dynamic nature of biphasic mixtures used in separation and purification strategies [25–27]. From a catalysis perspective, such efforts are of potentially high value in processes like metal scavenging during post-catalytic reaction clean-up [28].

An illustrative example of such analysis, revealing several key features of computer vision methodology in a chemistry context, was reported by Kapur's team and industrial collaborators in 2023 [25]. Here, the team set up a monochrome camera to capture grayscale video frames over time (1 FPS) and used these data to analyze layer heights in biphasic mixtures. The chemical composition of the mixtures was designed to represent common aqueous/organic liquid–liquid combinations, generating visibly separate layers over a range of time periods or, in extreme cases, generating stable emulsions that never settle into a visible bilayer. A key feature employed in the early steps of the algorithms included a low-pass filter to smooth (or blur) the raw grayscale image frame to remove high-frequency noise prior to more in-depth analysis. More specifically, the primary purpose of the low-pass finite impulse response (FIR) filter used is to allow low-frequency components of a signal to pass through while reducing the amplitude of the high-frequency components [29]. The FIR filter processes the input grayscale image by performing a convolutional operation similar to the one introduced in Section 2.2.3 (Figures 1.13 and 1.14). When you apply this filter to a grayscale image, each pixel in the output image is the average of the corresponding pixel in the input image and its neighbors. The output image appears smoother because the high-frequency components (sharp edges and noise) are reduced. Fine details and sharp transitions are blurred. While fine details are smoothed out, the overall structure and low-frequency components of the image (such as larger regions and gradual transitions) will be preserved.

With the use of the FIR filter, the team was then able to provide a rare example of using the first, second, and third derivatives of the smoothed grayscale image frame to help extract the sediment and cream phase fronts. These derivatives, with respect to pixel coordinates, provide important information about the changes in intensity values and highlight edges in the image via abrupt changes in intensity. High values of the first derivative, therefore, indicate the presence of edges. The magnitude of the gradient gives the strength of the edge, and the direction of the gradient indicates the orientation of the edge. The second derivative of a grayscale image measures the rate of change of the first derivative and provides information about the curvature of the intensity surface. Points where the second derivative changes sign indicate potential edges. These so-called zero-crossings are used in edge detection problems, as exemplified in Kapur's time-resolved detection of bilayers. The same second derivative can enhance the edges detected by the first derivative, making them more prominent.

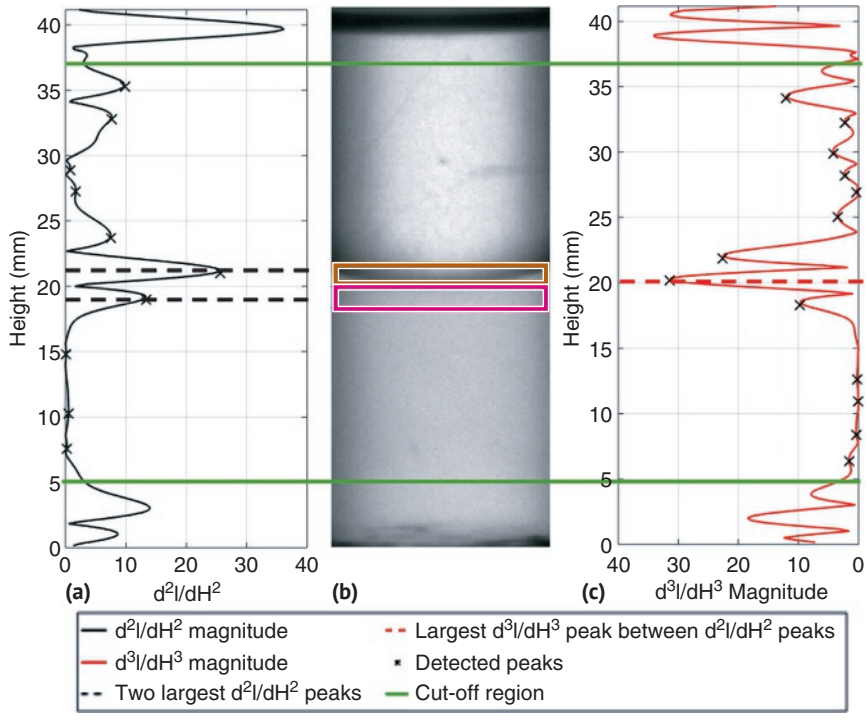

Figure 1.21 From Ref. [25], second (left) and third (right) derivative plots of grayscale gradient across the central image of a biphasic mixture. In the image, the bound boxes represent regions identified by the second derivative. The intersection of these areas is highlighted by a peak in the third derivative. [58]/Elsevier/CC BY 4.0.

The more rarely applied third derivative of the grayscale image measures the rate of change of the second derivative. While it is less commonly used in basic image processing tasks, it provides information about higher-order changes in intensity. Points where the third derivative changes sign can indicate inflection points in the intensity profile (Figure 1.21). In this case, Kapur was able to use the second derivative of the final video frame to locate positions near the interface within each of the two layers. The third derivative provided the reassuring location of the interface, between the positions found by the second derivatives of the top and bottom layers.

Each derivative order provides different levels of detail and can be used for various image processing tasks, from basic edge detection to more complex feature extraction and analysis. Consistent with earlier discussion on high-throughput methods, Kapur's team also made limited but promising attempts to use their method of monitoring phase separation dynamics in high-throughout developing a rig on which multiple biphasic samples could be mixed before being recorded using the monochrome camera chosen by the team. In relation to the introductory discussion on photography, it is worth noting that the monochrome camera is grayscale by its nature, without intermediate conversion of any RGB data to a linearized grayscale format. In other words, no Bayer (or other RGB) filter is applied before the incoming light is focused on the image sensor. Such cameras can be useful when higher sensitivity is needed, while color information is not.

1.3.2.2 Reaction Kinetics

From reaction monitoring [30] to solubility testing [31] and chemical networking [32], the use of video as an input data source for kinetic analysis via computer vision and ML is, at the time of writing, a growing trend.

In this author's team's efforts to develop time-resolved computer vision for reaction monitoring, including applications in catalysis, a necessary challenge in time series analysis presented itself. Namely, to understand the level of agreement between highly time-resolved imaging data from videos of reaction bulk and comparatively more sparse time series data collected from samples taken from the same reaction mixture for offline, more molecularly specific analyses (e.g., NMR, HPLC). In doing so, a ML approach was required such that any probable correlations between imaging and more established analytics could be determined without necessarily assuming whether any resulting model should be linear or otherwise.

To this end, mutual information (M.I.) analysis, based on Claude Shannon's pioneering quantification of information theory, became a natural choice [33, 34]. Here, entropy is used as the measure of the uncertainty associated with a random variable. The more uncertain we are, the more we lack information, where information is the capacity to reduce uncertainty. If a data source has a lowprobability value, the event carries more "surprisal" than when the source data has high probability. Overall, M.I. is a measure of the mutual dependence between two variables. It quantifies the amount of information obtained about one variable through the other, thus serving as a powerful tool for identifying correlations in time series datasets. Mathematically, M.I. I(X;Y) between two variables, X and Y is defined according to Eq. (1.19):

$$\begin{aligned} I(X;Y) &= -\sum_{x\in X}\sum_{y\in Y} P(x,y)\ .\ log_2\left[\frac{P(x,y)}{P(x)\ .\ P(y)}\right] \\ &= P(x,y)\ .\ \{\ log_2[P(x,y)] - log_2[P(x)\ .\ P(y)\ ]\} \end{aligned} \tag{1.19}$$

where $P(x,y)$ is the joint probability distribution of X and Y, and $P(x)$ and $P(y)$ are the marginal probability distributions of X and Y, respectively. A higher M.I. value indicates a stronger dependency between the variables, making it especially useful in analyzing the relationships and dependencies within multiple time series datasets. By capturing both linear and non-linear relationships, M.I. provide a more comprehensive measure compared to traditional correlation coefficients. Some non-mathematical intuition for the M.I. approach is provided in Figure 1.22.

One way to estimate marginal probability is the histogram approach (Figure 1.21a and Eq. 1.20). The marginal probabilities to be summed are then the frequency counts falling within each histogram bin divided by the total number of samples (i.e., elements in the time series array).

$$P(x_i) = \frac{no.\ elements\ in\ bin}{total\ no.\ elements\ in\ series} \tag{1.20}$$

If there are 10 histogram bins, the i ranges from 1 to 10. The process repeats for the number of supposedly independent features to be compared. After calculating the entropy for each feature, one can then look to understand the information available in one feature when another feature is given. This is where joint entropy (Eq. 1.21) comes in.

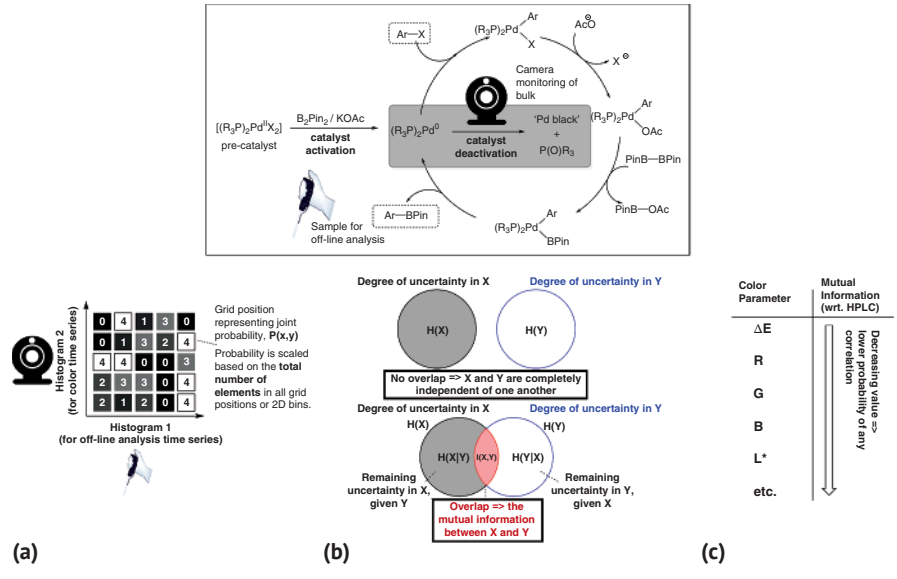

Figure 1.22 Top: A mechanistic representation of the Miyaura borylation, highlighting processes and analytes for which a combined monitoring approach employing cameras and HPLC analysis was employed. (a) A histogram approach to working out joint probabilities of occurrences from the independent computer vision and time series datasets. (b) A Venn diagrammatic overview of calculating M.I., I(X;Y). Part C: The practical output of a table showing a ranked order of color parameters that are decreasingly likely to correlate – linearly or otherwise – with the ground truth concentration data from HPLC.

The joint entropy of discrete random variables X and Y is the entropy of their pairing, H(X;Y). By the same chord, we now need to consider joint probability, $P(x_i, y_i)$, instead of marginal probability.

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} P(x,y) \; . \; log_2 [P(x,y)] \tag{1.21}$$

As a means of building intuition, (X,Y) could represent the position of a chess piece. X stores the row position, Y stores the column position. The entropy of the row of the piece and the entropy of the column of the piece come together to represent the entropy of the position of the piece. To estimate the joint probabilities, we can use 2D instead of 1D histograms. Each 1D histogram composing the 2D histogram can come from two "independent" features (Eq. 1.22).

$$P(x,y) = \frac{no.\ elements\ in\ 2D\ grid\ position}{total\ no.\ elements\ in\ the\ 2D\ grid} \tag{1.22}$$

Taking all of this toward an understanding of the M.I. held between two features or variables requires a Venn diagram visualization of the entropies thus far formulated (Figure 1.20). In reading these Venn diagrams:

- As overlap increases, the amount of shared information between X and Y increases. It does NOT mean there is any additional transfer of information between X and Y.

- When the M.I. is expressed using a *comma* between X and Y, it assumes the sizes of X and Y are the same. It also assumes we are only considering two features, X and Y. For example, X could be an array of HPLC samples over time. Y could be the subset of color data at the same time points as HPLC samples, making arrays X and Y the same size.
- The use of a semicolon is more general (e.g., for conditional M.I. involving more than two features). The semicolon denotes everything that is to be measured (left) given everything known (right).

Using this approach, it has been possible to rank order 10+ color parameters, collected and calculated with computer vision methods, in terms of their likely correlation with offline measures of product concentration by NMR and HPLC. Applications include the determination of reaction progress for the above-mentioned tracking of successful versus compromised Pd-catalyzed borylations [30], large-scale S_NAr reactions [10], and solid-phase peptide synthesis [35].

1.4 Summary and Conclusion

From a chemist's perspective, the field of computer vision is one of intriguing contradictions. On the one hand, it is a rapidly maturing field, with otherworldly applications developed for self-driving cars, factory robotics, drones and satellites, and cybersecurity. On the other hand, such sophisticated developments are surprising when they are compared to the relative infancy of computer vision applied to chemical analysis. This point is further emphasized when considering the near absence of any literature attending ML, computer vision, and catalysis.

In this chapter, we have considered selected illustrative examples of cameras being used to capture various visually characterizable chemical phenomena. Whether via single image or video, the availability of the singularly large datasets represented by arrays of pixels represents inordinate opportunities for the application and development of chemically intelligent ML applications.

Fueled by the wealth of transferable ML architectures and ever-increasing computing power available, it is envisaged that an increased number of video-focused computer vision applications will be realized for catalysis and computer vision. In the context of ML, the case studies shared show that, in the field of computer vision, much of the learning has as much to do with the machine (or camera) you choose as it does with the algorithms we more commonly associate with the term "machine learning."

References

1 Witzel, C., Racey, C., O'Regan, J.K. The most reasonable explanation of "the dress": implicit assumptions about illumination. *J Vis.* **2017** *17*, (2), 1 [Internet]. [cited 2024 Jun 10]; Available from: http://jov.arvojournals.org/article.aspx?https://doi.org/10.1167/17.2.1.

2 Krishna Prasad, K.M.M., Raheem, S., Vijayalekshmi, P., Kamala Sastri, C. Basic aspects and applications of tristimulus colorimetry. *Talanta*. **1996**, *43*, (8),1187–206.

3 Pascale, D. A review of RGB color spaces from xyY to R'G'B'. *Babel Color*. **2003**, 18, 136–152.

4 Specification of sRGB [Internet]. Specification of sRGB. Available from: https://www.color.org/chardata/rgb/srgb.pdf (accessed 3 Sept 2024)

5 Haralick, Robert M., Shanmugam, K., and Dinstein, I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern*. **1973** *SMC-3*, (6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

6 Alpern, M., Kitahara, K., Krantz, D.H. Classical tritanopia. *J Physiol*. [Internet]. **1983** *335*, (1), 655–681. Available from: https://physoc.onlinelibrary.wiley.com/doi/10.1113/jphysiol.1983.sp014557

7 Capitán-Vallvey, L.F., López-Ruiz, N., Martínez-Olmos, A., Erenas, M.M., Palma, A.J. Recent developments in computer vision-based analytical chemistry: a tutorial review. *Anal Chim Acta*. [Internet]. **2015**, *899*, 23–56. Available from: https://linkinghub.elsevier.com/retrieve/pii/S000326701501243X

8 RGB to grayscale – skimage 0.24.0 documentation [Internet]. Available from: https://scikit-image.org/docs/stable/auto_examples/color_exposure/plot_rgb_to_gray.html#id2 (accessed 3 Sept 2024)

9 Davis, Larry S. and Johns, Steven A. and Aggarwal, J. K., Texture Analysis Using Generalized Co-Occurrence Matrices. *IEEE Trans Pattern Anal Mach Intell*. **1979**, PAMI-1, (3), 251–259. https://doi.org/10.1109/TPAMI.1979.4766921.

10 Barrington, H., Dickinson, A., McGuire, J., Yan, C., Reid, M. Computer vision for kinetic analysis of lab- and process-scale mixing phenomena. *Org Process Res Dev*. **2022**, *26*, (11), 3073–3088.

11 Davis, L.S., Johns, S.A., Aggarwal, J.K. Texture analysis using generalized co-occurrence matrices. *IEEE Trans Pattern Anal Mach Intell*. **1979**, PAMI-1(3), 251–259.

12 Perez Alvarado, F.A., Hussein, M.A., Becker, T.A. Vision system for surface homogeneity analysis of dough based on the grey level co-occurrence matrix (GLCM) for optimum kneading time prediction. *J Food Process Eng*. [Internet]. **2016**, *39*, (2), 166–177. [cited 2021 Sep 20]; Available from: https://onlinelibrary.wiley.com/doi/10.1111/jfpe.12209

13 Gosselin, R., Duchesne, C., Rodrigue, D. On the characterization of polymer powders mixing dynamics by texture analysis. *Powder Technol*. [Internet]. **2008**, 183, (2), 177–188. [cited 2021 Sep 29]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0032591007003646

14 Jing, J., Liu, S., Wang, G., Zhang, W., Sun, C. Recent advances on image edge detection: a comprehensive review. *Neurocomputing*. **2022**, *503*, 259–271.

15 Baillargeon, P., Scampavia, L., Einsteder, R., Hodder, P. Monitoring of HTS compound library quality via a high-resolution image acquisition and processing instrument. *J Lab Autom*. **2011**, *16*, (3), 197–203.

16 Sun, X., Boulgakov, A.A., Smith, L.N., Metola, P., Marcotte, E.M., Anslyn, E.V. Photography coupled with self-propagating chemical cascades: Differentiation

and quantitation of g-and v-nerve agent mimics via chromaticity. *ACS Cent. Sci*. **2018**, *4,* (7), 854–861. [cited 2024 Jun 10]; Available from: https://github.com/

17 Eppel, S., Xu, H., Bismuth, M., Aspuru-Guzik, A. computer vision for recognition of materials and vessels in chemistry lab settings and the vector-labpics data set. *ACS Cent Sci*. [Internet]. **2020**, *6*, (10), 1743–1752. [cited 2020 Sep 18]; Available from: https://pubs.acs.org/doi/10.1021/acscentsci.0c00460

18 Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. [Internet]. **2015**, *115*, (3), 211–252. Available from: http://link.springer.com/10.1007/s11263-015-0816-y

19 He, K., Gkioxari, G., Dollár, P., Girshick, R. Mask R-CNN. In: 2017 *IEEE Int. Conf. Comput. Vis (ICCV)*. **2017**, 2980–2988.

20 Tu, J., Li, H., Yan, X., Ren, M., Chen, Y., Liang, M., Bitar, E., Yumer, E., and Urtasun, R. Exploring adversarial robustness of multi-sensor perception systems in self driving. In: *Proceedings of the 5th Conference on Robot Learning* (eds. A. Faust D. Hsu G. Neumann), [Internet]. PMLR; **2022**, 1013–1024. (*Proceedings of Machine Learning Research*; vol. 164). Available from: https://proceedings.mlr.press/v164/tu22a.html

21 Rodrigues, N.V.N., Abramo, L.R., Hirata, N.S.T. The information of attribute uncertainties: what convolutional neural networks can learn about errors in input data. *Mach Learn Sci Technol*. [Internet]. **2023**, *4*, (4), 45019. Available from: https://dx.doi.org/10.1088/2632-2153/ad0285

22 Grudpan, K., Kolev, S.D., Lapanantnopakhun, S., McKelvie, I.D., Wongwilai, W. Applications of everyday IT and communications devices in modern analytical chemistry: a review. *Talanta*. **2015**, *136*, 84–94.

23 Ley, S.V., Ingham, R.J., O'Brien, M., Browne, D.L. Camera-enabled techniques for organic synthesis. *Beilstein J Org Chem*. **2013**, *9*, 1051–1072.

24 Apyari, V.V., Gorbunova, M.V., Isachenko, A.I., Dmitrienko, S.G., Zolotov, Y.A. Use of household color-recording devices in quantitative chemical analysis. *J Anal Chem*. **2017**, *72*, (11), 1127–1137.

25 Daglish, J., Blacker, A.J., de Boer, G., Crampton, A., Hose, D.R.J., Parsons, A.R., Kapur, N. Determining phase separation dynamics with an automated image processing algorithm. *Org Process Res Dev*. [Internet]. **2023**, *27*, (4), 627–639. [cited 2024 Jun 4]; Available from: https://pubs.acs.org/doi/full/10.1021/acs.oprd.2c00357

26 Duffield, S., Da Vià, L., Bellman, A.C., Chiti, F. Automated high-throughput partition coefficient determination with image analysis for rapid reaction workup process development and modeling. *Org Process Res Dev*. [Internet]. **2021**, *25*, (12), 2738–2746. [cited 2022 Apr 1]; Available from: https://pubs.acs.org/doi/abs/10.1021/acs.oprd.1c00343

27 El-Khawaldeh, R., Guy, M., Bork, F., Taherimakhsousi, N., Jones, K.N., Hawkins, J.M., Han, L., Pritchard, R.P., Cole, B.A., Monfette, S. and Hein, J.E. Keeping an "eye" on the experiment: computer vision for real-time monitoring and control. *Chem Sci*. [Internet]. **2024** *15*, (4), 1271–1282. [cited 2024, Jun 10]; Available from: https://pubs.rsc.org/en/content/articlehtml/2024/sc/d3sc05491h

28 Economidou, M., Mistry, N., Wheelhouse, K.M.P., Lindsay, D.M. Palladium extraction following metal-catalyzed reactions: Recent advances and applications in the pharmaceutical industry. *Org Process Res Dev*. [Internet]. **2023**, *27*, (9), 1585–1615. [cited 2024 Jun 10]; Available from: https://pubs.acs.org/doi/full/10.1021/acs.oprd.3c00210

29 Sundararajan, D. Finite impulse response filters. In: *Discrete Wavelet Transform: A Signal Processing Approach*. **2015**, 63–70.

30 Yan, C., Cowie, M., Howcutt, C., Wheelhouse, K.M.P., Hodnett, N.S., Kollie, M., Gildea, M., Goodfellow, M.H., Reid, M. Computer vision for non-contact monitoring of catalyst degradation and product formation kinetics. *Chem Sci*. **2023**, *14*, (20), 5323–5331.

31 Shiri, P., Lai, V., Zepel, T., Griffin, D., Reifman, J., Clark, S., Grunert, S., Yunker, L.P., Steiner, S., Situ, H. and Yang, F. Automated solubility screening platform using computer vision. *iScience*. **2021**, *24*, (3), 102176.

32 Caramelli, D., Salley, D., Henson, A., Camarasa, G.A., Sharabi, S., Keenan, G., Cronin, L. Networking chemical robots for reaction multitasking. *Nat Commun*. [Internet]. **2018**, *9*, (1), 1–10. [cited 2021 Mar 30]; Available from: www.nature.com/naturecommunications

33 Dionisio, A., Menezes, R., Mendes, D.A. Mutual information: A measure of dependency for nonlinear time series. *Phys A: Stat Mech Appl*. [Internet]. **2004**, *344*, (1–2), 326–329. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0378437104009598

34 Shannon, C.E. A mathematical theory of communication. *Bell Syst Tech J*. [Internet]. **1948**, *27*, (3), 379–423. [cited 2022 May 10]; Available from: https://ieeexplore.ieee.org/document/6773024

35 Yan, C., Fyfe, C., Minty, L., Barrington, H., Jamieson, C., Reid, M. Computer vision as a new paradigm for monitoring of solution and solid phase peptide synthesis. *Chem Sci*. [Internet]. **2023**, *14*, (42), 11872–11880. Available from: https://xlink.rsc.org/?DOI=D3SC01383A